



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 112389

TO: Phuong Bui
Location: REM-2A15
Art Unit: 1638
Jan 22, 2004

Case Serial Number: 09831233

From: P. Sheppard
Location: Remsen Building
Phone: (571) 272-2529

sheppard@uspto.gov

Search Notes

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 19:35:55 ; Search time 7195 Seconds
(without alignments)
11121.505 Million cell updates/sec

Title: US-09-831-233A-1

Perfect score: 1956

Sequence: 1 gcacgaaggaactggca.....aaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.ey.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.ay.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	928.4	47.5	1996	8	AK118498 Arabidops
2	927.4	47.4	1779	8	BT005923 Arabidops
3	902.2	46.1	1994	8	AB006210 Arabidops
4	359.6	18.4	44308	8	AL033388 S.pombe c
5	336	17.2	2429	8	SPBC887
6	336	17.2	2440	8	SCYBR248C
7	303.2	15.5	87435	8	NC103E1
8	266.6	13.6	4351	8	AF159463 Emericell
9	234.8	12.0	110000	2	Continuation (2 of
10	224.2	11.5	975	11	AL402350 T3 end of
11	198.2	10.1	2306	8	AF290177 Candida a
12	150.8	7.7	4581	8	AB016783 Arabidops
13	129.6	6.6	195583	8	AL134234 Oryza sat
14	127.8	6.5	280050	1	AL591975 Listeria
15	127.8	6.5	349980	6	AX641665 Sequence
16	127.8	6.5	349980	6	AX641666 Sequence
17	126.4	6.5	11440	1	AE014962 Streptoco
18	126.4	6.5	128294	8	AC135225 Oryza sat
19	126.4	6.5	137580	8	AF377947 Oryza sat
20	125.4	6.4	249050	1	AL596165 Listeria
21	125.4	6.4	349980	6	AX417038 Sequence
22	125.4	6.4	349980	6	AX417041 Sequence
23	125.4	6.4	349980	6	AX417042 Sequence
24	117.2	6.0	4584	6	AX416785 Sequence
25	116	5.9	12075	1	AE013160 Thermoana
26	112.2	5.7	11308	1	AE007609 Clostridi
27	107	5.5	15451	1	AE000677 Aquifex a
28	106.2	5.4	128574	8	ATFL0M23
29	106.2	5.4	198372	8	ATCHRIV66
30	102.8	5.3	12027	1	AE000897
31	101.2	5.2	12634	1	AE000912 Methanoba
32	100.2	5.1	47739	1	AF017113 Bacillus
33	100.2	5.1	209510	1	BSUB0018
34	98.6	5.0	11141	1	U67493
35	98.6	5.0	110000	6	AR271569_10
36	98.6	5.0	110000	6	AR271569_11
37	96	4.9	8313	1	SSU82227_11
38	96	4.9	13528	1	AE006689
39	96	4.9	281244	1	SSU18930
40	96	4.9	300150	1	AP004594
41	95	4.9	10386	1	AE010714
42	94	4.8	306050	1	EX321858 Nitrosomo
43	93.6	4.8	10515	1	AE012824 Chlorobiu
44	92.4	4.7	10186	1	AE010316 Methanopy
45	92.2	4.7	3300	1	AF150930 Thermoana

ALIGNMENTS

RESULT 1
AK118498
LOCUS
DEFINITION
Arabidopsis thaliana At4G26900 mRNA for putative glutamine
amidotransferase/cyclase, complete cds, clone: RAFL19-73-E07.
ACCESSION
AK118498
VERSION
AK118498.1 GI:26452023
KEYWORDS
FLI CDNA; CAP trapper.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1
AUTHORS
Seki,M., Iida,K., Satou,M., Sakurai,T., Akiyama,K., Ishida,J.,

Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayaehizaki, Y. and Shinozaki, K.
Arabidopsis thaliana full-length cDNA
Published Only in Database (2002)
2 (bases 1 to 1996)
Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayaehizaki, Y. and Shinozaki, K.
Direct Submission
Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: msek@gs.riken.go.jp, URL: <http://pfigweb.gsc.riken.go.jp>, Tel: 81-45-503-9625, Fax: 81-45-503-9586)
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector.
Please visit our web site (<http://pfigweb.gsc.riken.go.jp/>) for further details.

Location/Qualifiers
1..1996
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
/clone="RAP19-73-E07"
/note="common name: thale cress"
1..1996
/gene="At4g26900/F10M23_240"
33..1811
/gene="At4g26900/F10M23_240"
/codon_start=1
/product="putative glutamine amidotransferase/cyclase"
/protein_id="BAC433102.1"
/db_xref="GI:26452024"
/translation="MEATAAPFSSIVSSRNQFSSSSIRASSPASIFLSOKSIGNVR
KFKSPRLSVRASSTDSVVTLLDYAGNVRISRNALRHGFGSIKDVQTPGDLNADR
LIPGVGAFAPAMDVLRNTGMAELCKYIENDRPFLGICLGLQLLPDSSSENGPVGKL
GVIPIVGRPDASGIRVPHIGNVALQVKDSEILDVGNRVVVFVHSYRAIPSDENK
DMISSTGSPFSSIRRNHVAQVHPKSESGEVLVLRRLPHKLPATOKMEBK
ASKAKRVIACLDVRNDGLVTVKGQIDVREQSNENVRNLKRFVDLAGQYKDG
ADEISFLNITGFRDPLPMIQLVRLQTSKNVPVLTGVGGIRDFTDASGRYISLE
VAAEYFRSGADKISIGSDVSAABEEFKSGVTKSLEQISRVYGNQAVVVIDPRR
VYVNHDPDVPYKIVRTNPNPENGVEYARYQCTVSGGREGRPIGAFELAKAVEELGAE
ILLNCIDCGQGFIDLVLKILSDSVGIPVIASSGAGTDPHFSEVPEKTNASALAA
GIFRKEVPISQVKEHLQERIEVRI"
BASE COUNT 555 a 386 c 491 g 564 t

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Query Match 47.5%; Score 928.4; DB 8; Length 1996;
Best Local Similarity 73.8%; Pred. No. 2.3e-198;
Matches 1210; Conservative 0; Mismatches 421; Indels 9; Gaps 2;
Qy 43 CAGAAATCGCCGCCCTTGGCTGGTGCGAGCGGAGATTCTGTGTGATCTTTACTTTGATTA 102
Db 179 CAGAAGCTCTCCGCTCCGCGCATCTTCTACCTCAGATTCTGTGTGACTTGTCTTGACTA 238
Qy 103 CGGTCTGGGAATCTTAGGAGTGTGAGGACGCCATCGCACACTTCGATTTGATATCAA 162
Db 239 CGGAGCTGGAATGTTCGAGAGCATCCGCAATGCTCTTCGTCTATCTCGGCTTCAGCATCA 298
Qy 163 AGATGTGCAAAAGCAGAGGATATCTTAAATGCTAAGCGCCCTTATCTTCTCGCGTTGG 222
Db 299 AGAGCTTCAACGCGGAGACATCTCAATGCTGATCGACTCATATTTCCAGCGCTGG 358
Qy 223 GGCCTTTGACCTGCAATGGATGTTCTTATTTGTAAGAGGGCTGCTGAAACACTCTCTGAC 282
Db 359 TGCCTTTGACCCGCCATGGATGACTTAAACAGAACTGGATGGCTGAAAGCTTTGTGCAA 418

Qy 283 TTACATTCAGATGATGACGACCTTTCTCTGGGTATATGCTGGGATTCGAGCTACTCTTTGA 342
Db 419 ATATATTTAGATGACGCTCCATTTCTAGGCATATGCTTGGTCTACAACACTACTTTTGA 478
Qy 343 GTCAGTGAAGAAAATGGTCCCAATTCAGGCTCTGGCTTGAATTCCTGGAGCGGTTGGCG 402
Db 479 TTTAGTGAAGAGATGGACCACTCAAGGCTCTGGTGTGATACCGGGAATAGTTGGACG 538
Qy 403 TTTTGAATCATCAATGGTCTTAAAGGTGACCATATTTGGATGCGATGCGCTTGGATATAA 462
Db 539 CTTTGTATGCTTACGCTGGGTATTAAGAGTACCCACATTTGGCTGGAATGCTTTTCAAGTGG 598
Qy 463 GGAAGGCTCAGCAATTTTATAGATGATGCTGGGGAATCAACATGCTGATTTTCTTCACTCAT 522
Db 599 GAAGGATTCGAATTTTGGATGATGCTGGAACCGCTCATGCTATTTTGTTCATTCGTA 658
Qy 523 TCAGAGCAATGTC--GAGGACAAAGAGTGGATTTTCATCTACATGACGATGATGTTGA 579
Db 659 CAGGGCAATTCATCAGATGAAATTAAGGACTGGATTTCTGCTACCTGTATATATGTTGA 718
Qy 580 CGATTTTATGTCATTCATTCAGAGGGAATGTTTCATGCTGAGTCCAAATTTTCATCCCGAG 639
Db 719 ATCAATTTATATCTTCCATTAAGAGGGAATGTCATGCTGAGTTCAATTTCCATCTCGAAA 778
Qy 640 GAGTGGAGGTGTTGGACTTTTCCATATTTGAGAGATTTTTCGAATGCTGATTTCTTTTAA 699
Db 779 GAGCGGGAGGTGGGCTTCTGTTTAAAGAGGTTCTTGCATCCAAAATTAACCT---- 833
Qy 700 CAAAGACAGAGCCCAATGAATGGAAGGCTTCTAAACTTTGCAAGAGAGTAAATTTGCTTG 759
Db 834 -GCAACACAGAGCCCAATGGAAGGAAAGGCTCAAAACTTTGCAAGAGGAGTAAATTTGCTTG 892
Qy 760 CTTTGTATGTCAGGCGCAATGATTAATGGGATCTTGTGTTAAACCAAGGGAGCAATATGA 819
Db 893 TCTTGTATGTCAGGCGCAATGATTAAGGAGATCTCGTAGTTACTTAAAGGGGATCAGTATGA 952
Qy 820 TGTGAGAGACGTACAGAGAGATGAGGTGCAAAACCTTTGGCAAGCTCTTGAACCTTGC 879
Db 953 TGTGAGAGACCAATCTAATGAAAACAGGTTCCAAACCTTTGGCAACCTTGTGATTTGGC 1012
Qy 880 TGGCGAGTATATTTAGACGCTCTGATAGGTCAGCTCTCTTAAACATTTACTGTTTCCG 939
Db 1013 TGGCGAGTATTAACAAGATGGTCAGATGAGATTAGCTTTTAAACATACTCGATTTCCG 1072
Qy 940 GGAATTCCTCTAGGCGATCTACCCATGCTACAGGTTCTTCAACGCGCTCTGAAACGT 999
Db 1073 CGATTTCTCTAGGCGATTTGCGGATGATTCAGGTTGTTGAGGCGACATCAAGAATGT 1132
Qy 1000 TTTTGTGCCATTAACCTGTCCGGGCTGGCATCAGGATTTTCTGATGCAAAATGGAAGTA 1059
Db 1133 CTTTGTACCACTAATCTGTTGGAGTGGTATTAGAGACTTTACAGATGCTAGTGGCAGTA 1192
Qy 1060 TTATTTCTAGTCTAGAAAGTGGCTTCAGAGTATTTTCAGATCGGGCGCGGATTAAGTTTCAT 1119
Db 1193 CTATTTCTAGCTTGAAGTTGCTGCTGATATTTTCAGATCCGGTGTCTGATTAAGATCTCCAT 1252
Qy 1120 CGGAGTGTGAGTTTACACTGCTGAGGATATATTAACCCGGAGTGAAGCAGGAAA 1179
Db 1253 AGGAAGTGAAGCTGTTTCTGCTGAGAGGAGTTTCATAAAATCAGGGGTGAAGCAGGAAA 1312
Qy 1180 GAGCAGCATAGAGCAGATATCTACAGTATATGTTAAACCCAGGCGAGTGGTTGTAAGCATTTGA 1239
Db 1313 GAGTAGTTTGAACAGATCTCAGAGTTTATGGAATCAGGAGTGGTTGTAAGTATTTGA 1372
Qy 1240 TCTTCGCGAGTTTACTTTGAGAAAACCCGATGAAGTAGAATTTTAAAGCCATCAAGTAAG 1299
Db 1373 TCTTCGTAGAGTTTATGTGAAACCATCCGATGATGTGCCATACAAAGTCTCAGAGTAAC 1432
Qy 1300 CCATCCAGTTCAAAACGGTGAAGAAATATGCTGATACGTGCTGCTGTTTAAATGTTGACG 1359
Db 1433 TAAATCCAGGCGCCAAATGGAAGAAATATGCCCCGTATCAGTGAACGCTAGTGGAGACG 1492
Qy 1360 AGAAGGAGAGCCCATCGGAGCTTTATGAATAGCTTAAGGCTGTTTGAAGAACTTTGGAGCTGG 1419

627 CAGGGCCATCCATCAGATGAAATAAGGACTGGATTTCGTCTACCTGTAATTATGCTGA 686
Db
580 CGATTTTATTCATCCATTGAGAGGGAATGTTTCATGCAAGTCAATTTTCATCCCGAA 639
Qy
687 ATCATTTATATCTTCCATAAGAGGGGAATGTCATGCAAGTCCAAATTCATCTCGAANA 746
Db
640 GAGTGGAGGTGTTGCACTTTCCATATTGAGAAGATTTTGAATGCTGATTCCTTTAAACA 699
Qy
747 GAGCGGGAGGTGGGCTTTCTGTTTAAAGAGGTTCTTGATCCAAATTAACCT----- 801
Db
700 CAAAGACAGAGCAATGAATGAAAGGCTTTAAACTCTGCAAGAGAGTAATGCTTG 759
Qy
802 -GCAACACAGAGCAATGGAAGGAAAGGCTCAAAACTTTCGCAAGAGGTTGATTGCTTG 860
Db
760 CTTTGTGAGGCAATGATTAATGCGGATCTTGTGTAACCAAGGAGACCAATATGA 819
Qy
861 TCTTGTGATGAGGCAATGATAAGGAGATCTGTAAGTACTTAAGGGGATCAGTATGA 920
Db
820 TGTGAGAGAGCTACAGAGAGATGAGGTCAGAAACCTTGGCAAGCTGTTGAACCTTGC 879
Qy
921 TGTGAGAGAGCAATCTAATGAAACGAGGTTGCAAACTTGGCAAACTGTTGATTTGGC 980
Db
880 TGGCAGTATTAATTAGACGGTGTGATGAGGTGAGCTTCTTAACATTAATGCTGTTCCG 939
Qy
981 TGGCAGTATTAACAAGATGCTGAGATGAGATTAAGCTTTTAAACATAACTGGATTCCG 1040
Db
940 GGACTTCTCTAGCGGATCTACCATGCTACAGGTCTTGCACCGCGATCTGAAACGT 999
Qy
1041 CGATTTTCTCTAGGGATTTGCGGATGATTCAGGTGTTGAGGAGACATCAAGAATGT 1100
Db
1000 TTTTGTGCAATTAAGTGGGGTGGCATGAGGATTTTACTGATCAATGGAAGGTA 1059
Qy
1101 CTTTGTACCACTAAGTGGAGGTGATTAAGAGATCTTACAGATCTAGTGGCAGGTA 1160
Db
1060 TTATTTAGTCTAGAGTGGCTTCAGAGTATTTTCAGATCGGGCCCGCATAGGTTTCGAT 1119
Qy
1161 CTATTTAGCTTGGAGTTGCTGCTGAGTATTTTCAGATCCGCTGCTGATAGATCTCCAT 1220
Db
1120 CGAAGTGTATGCTGCTGAGGAAATATATTAATAACCGGAGTGAAGACAGGAA 1179
Qy
1221 AGAAGTGAACCTGCTTCTGCTGAGAGGAGTTCATAAAATCAGGGGTGAAGACAGGAA 1280
Db
1180 GAGCAGCATAGACAGATATCTACAGTATATGTAACAGGAGTGTGTAAGCATTTGA 1239
Qy
1281 GAGTAGTTTGAACAGATATCCAGATTTTGAAGATCAGGAGTGTGTTAGTATTTGA 1340
Db
1240 TCCTCGCGAGTTTACTTTGAGAAACCCGATGAAAGTAAATTTAAAGCCATCAAGTAAG 1299
Qy
1341 TCCTCGTAGAGTTTATGTGAACCATCCGATGATGTCATACAAAGTCAATCAGAGTAAC 1400
Db
1300 CCATCCAGGTCCAAACGGTGAAGGAATATGCTGATGATGAGTGTGTAAGTGAAGC 1359
Qy
1401 TAAATCAGGCGCCAAATGAGAGGAATATGCCCGGTATCAGTGACGCTGAGTGGAGAGC 1460
Db
1360 -AGAGGAGGACCCATCGAGCTTATGACTAGCTTAAGCTGTTTCAGAGCACTTGGAGCTGG 1419
Qy
1461 AGAAGGTGACCTATGAGAGCAATTTGAGCTTGGCAAGCGGTTGAAGATTTAGTGGCGG 1520
Db
1420 AGAATATTTTGAACCTGATTTGATGAGTCAAGGAAAGGATTCGATATAGATCT 1479
Qy
1521 TGAATATTTGAACCTGATAGATGTTGATGTTCAAGGGAAGGATTCGACATAGACTT 1580
Db
1480 GATCAAGCTAATATCCGATGCTGGAATATTCCTGTTATCGCAAGCAGCGGTGAGAGT 1539
Qy
1581 AGTAAAGCTATCTCAGATTCAGTAGGATACCGGTGATCGCAAGCAGTGGAGAGTAC 1640
Db
1540 CGCTGATCAGCTTCTCCGAGTCTTTTAAAGAACCAAGCATCTGCTGCCCTTGGAGCTGG 1599
Qy
1641 TCCCGACCACTTTTCCGAGGTGTTTGAAGAACCAAGCATCTGCCCGCTTGTGCTCGG 1700
Db
1600 CATTTTCCATCGCAAGAGGTTTCAATTTAAGGCTGTTAAAGAGACATTTGTTGAAGGAGG 1659
Qy
1701 CATTTTCCACGGAAAGAGGTACCAATCCATCTGTGAAGAGGACATTTACAAGAGGAGCG 1760
Db

Qy 1660 GATTGAAGTTAGATTGTAA 1678
Db 1761 CATAGAAGTCAGATCTGA 1779

RESULT 3
AB006210 1994 bp mRNA linear PLN 05-FEB-1999
LOCUS Arabidopsis thaliana mRNA for glutamine amidotransferase/cyclase,
DEFINITION complete cds.
ACCESSION AB006210
VERSION AB006210.1 GI:3219163
KEYWORDS glutamine amidotransferase/cyclase.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsia.

REFERENCE 1 (sites)
AUTHORS Fujimori, K. and Ohta, D.
TITLE An Arabidopsis cDNA encoding a bifunctional glutamine
amidotransferase/cyclase suppresses the histidine auxotrophy of a
Saccharomyces cerevisiae his7 mutant
PEBS Lett. 428 (3), 229-234 (1998)

JOURNAL 98316711
MEDLINE 9654139
PUBMED 9654139

REFERENCE 2 (bases 1 to 1994)
AUTHORS Fujimori, K.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-1997) Ko Fujimori, Osaka Bioscience Institute,
Department of Molecular Behavioral Biology; 6-2-4 Furuedai, Suita,
Osaka 565-0874, Japan (E-mail: fujimori@obi.or.jp,
Tel: +81-6-872-4851, Fax: +81-6-872-2841)

FEATURES
Location/Qualifiers
1..1994
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
6..1787
/codon_start=1
/product="glutamine amidotransferase/cyclase"
/protein_id="BAA28783.1"
/db_xref="GI:3219164"
/translation="MEATAAPSSIVSRQNSRSSSSIRASSPASLFLSQKSGTGNVR
KFKSPRLSVRASSTSDSVVTLDDYAGNVRIRNALRHGFISIKDVQTPGDLNADR
LIPGVPFAPAMDVLNRTGMAEALCYIENDRPFLGICLGLQLFDSSEQNPVKGL
GVTPGVGRFASAGIRVPHIGWALQVGDSEILDVGNRHVYFVHSYRAIPSDENK
DMISCTNYGESFISIRGNHVAQVHPEKSEVGLSVLRFLPKLPATOKPMGK
ASKLKRVTACLDVRTNDKDLVTVKDDQDVREOSNENEVNLKGPDLAGQVYKDG
ADEISFLNITGFRDFPLGDLPMIQLVLRQSKNVFVLTGGRDFTDASGRYSSLE
VAAEYFRSGADKMSIGSDAVFAABEFIKSGVTGKSSLSQISKVYGNQAVVVIDPRR
VYVNHDPDVPYKIVRVNTPGNEEYATQCTVSGQGEGRPIGAFELAKAVEILGAGE
ILLNCINCQDQKGFIDLVKLISDSVGIPTVIASSAGTDPHFSEVFEDKRICRACC
RHPPEPGYQSQSVKEHLEQERIEVRI"
1994
poly_a_site
/note="17 a nucleotides"

BASE COUNT 557 a 385 c 487 g 565 t
ORIGIN

Query Match 46.1%; Score 902.2; DB 8; Length 1994;
Best Local Similarity 73.5%; Pred. No. 1.8e-192;
Matches 1208; Conservative 0; Mismatches 423; Indels 12; Gaps 4;
Qy 43 CAGAAATTCGCGCGCTTCGCTGGTGCAGCGGAGATTCGTGGTGACTTTACTTGATTA 102
Db 152 CAGAAGCTCTCGTCGCGCATCTTCTACCTCAGATTCCTGTGACTTTGCTGACTA 211
Qy 103 CGGTGCTGGAATGTTAGAGGTGTAGAGAACGCCATCCGACACTTGGATTTGATCAA 162
Db 212 CGGAGCTGGAATGTTTCGGAGCATCCGCAATGCTCTTCGTGTCATCTCGGCTTCAGCATCA 271

QY 163 AGATGTGCAAAAGCCAGAGGATATCTAAATGCTAAGGCGCTTATCTTTCTGCGGTGG 222
 Db 272 AGAGTTTCAAAAGCGCGGAGACATTTGCAATGCTGATCGACTCATATTTTCCAGGCGTTGG 331
 QY 223 GGCCTTTGCACTTCAATGATGTTCTTATTTGTAAGGGCTGCTGAAGCACTCTGTAC 282
 Db 332 GCCTTTTGCACCCCGCATGATGATCTTAACAGAACTGGGATGGCTGAAGCTTTGTCAA 391
 QY 283 TTACATTTAGAAATGATCGACCTTTCTTGGGTATATGCTCGGATGTCAGCTACTCTTTGA 342
 Db 392 ATATATTTAGAAATGACCGTCCATTTCTAGGATATGCTTGGTCTACAACTACTTTTGA 451
 QY 343 GTCAAGTGAAGAAATGGTCAATTTCAAGGTCTTGGCTTGAATCTCTGGAAGGGTGGCG 402
 Db 452 TTCTAGTGAACAGAAATGCAACAGTCAAGGTCTTGGTGTATACCGGGAATAGTTGGACG 511
 QY 403 TTTTGAATCATCAATGGTTTAAAGGTGACACATATTTGGATGGCATGCCCTTGGATATAA 462
 Db 512 CTITGATGCTTCACTGGTATAGAGTACCCACATTTGGCTGGGAATGCTTTGCAAGTTGG 571
 QY 463 GGAAGGCTCAGCAATTTAGATGATGTGGGGAATCAACATGTGTATTTTGTCTCACTATA 522
 Db 572 GAAGATTTCTGAATTTTGGATGATGTGGAAACCGTCACTGTCTATTTTGTCTATTCGTA 631
 QY 523 TCGAGCCAATGCC--GAGGACAAAGAGTGGATTTCTATCATGCACTATGTTGA 579
 Db 632 CAGGGCCATTCATCAGATGAATAAGGACTGGATTTCTGCTACCTGTAAATATGTTGA 691
 QY 580 CGATTTTATTCATCCATTCAGAAGGGAATGTTTCATGCACTCCAATTTTCATCCCGAA 639
 Db 692 ATCATTTATATCTTCCATNAAGAGGGGGAATGTGCATGCGATTCATTTCCATCCTGAAA 751
 QY 640 GAGTGGAGGTGTGGACTTTTCCATATTTGAGAAGATTTTGAATGCTGATTTCTTTTACAA 699
 Db 752 GAGCGGGAGGTGGGGCTTTCTGTTTAAAGAGTCTTGCATCCAAAATTAACCT-- 806
 QY 700 CAAAGACAGAACCAATGAATGAAGGCTTTTAAACTTTCGAAAGAGATTAATGCTTG 759
 Db 807 -GCAACAGAGAACCAATGGAAGGAGGCTTCAAACTTTCGAAAGAGGTGATTTGCTTG 865
 QY 760 CTTTGTGTAGGCAATGATATGAGGATCTTGTGTAAACAGGAGACCAATATGA 819
 Db 866 TCTTGTGTAGGAGCAATGATAAGGAGATCTGTATGTTACTTAAAGGGATCAGTATGA 925
 QY 820 TGTGAGAGACGTACAGAGAGATGAGTGCAGAACTTGGCAAGCTTGTGAACCTTGC 879
 Db 926 TGTGAGAGAGCAATCTAATGAACAGAGTTTTCGAAACCTTGGCAACCTGTTGATTTGGC 985
 QY 880 TGGGCACTATTTATAGACGCTGCTGATGAGGTGAGCTTCTTAAACATTAAGTTTCCG 939
 Db 986 TGGGCACTATTTACAAAGATGGTGCAGATGAGATTTAGCTTTTAAACATTAAGTTTCCG 1045
 QY 940 GGACTTCCCTTAGGCGATCTACCCATGCTACAGGTCTTGAACCGCGCATCTGAAGAGT 999
 Db 1046 CGATTTTCTCTAGGGAATTTGCGGATGATTCAGGTGTTGAGGCGAGACATCAAGAGATGT 1105
 QY 1000 TTTTGTCCATTAATCTGCGGGGTGCAATCAGGATTTTACTGATGCAATGGAAGTA 1059
 Db 1106 CTTTGTACCACTAACTCTGAGGTGATTTAGAGACTTTTACAGATGCTAGTGGCAGGTA 1165
 QY 1060 TTATTTCTAGTCTAGAGTGGCTTCAGAGTATTTTTCAGATCGGGCGCCGATGAAGTTTCAT 1119
 Db 1166 CTATTTCTAGCTTGAAGTTGCTGCTGAGTATTTTCAGATCCGCTGCTGATGAAGTTCAT 1225
 QY 1120 CGGAAGTGTAGCTAGTTTACACTGCTGAGGGAATATATTTAAACCGGAGTGAAGACAGAAA 1179
 Db 1226 AGGAAGTGAAGCTGCTTTTGTGTCAGAGGAGTTTCAATAAATCAGGGGTGAAGACAGGAAA 1285
 QY 1180 GAGCAGCATAGACAGATATCTACAGTATATGTTAAACCGGAGTGGTCTTGAAGCATTTGA 1239
 Db 1286 GAGTAGTTTGAACAGATATCCAGATTTTATGAAATCAGGCACTGGTGTGAAGTATTTGA 1345
 QY 1240 TCCTCGCGGAGTTTACTTTGAGAAAACCCGATGAAGTAGAATTTTAAAGCCATCAAGTAAG 1299

Db 1346 TCCTCTAGAGTTTATGTGAACCATCCGATGATGTGCCATACAAAAGTCATCAGAGTAAC 1405
 QY 1300 CCATCCAGGTCCAAACGGTGAGGAATATGCTCGTATCAGTGCACCTGTTTAAATGTTGAGC 1359
 Db 1406 TAATCCAGGCCCAATGGAAGAAATATGCTCGTATCAGTGCACCGTCTAGTGGAGGACA 1465
 QY 1360 AGAAGGAGACCCATCCGAGCTTTATGAACCTAGCTAAGGCTGTTGAGAACTTCGAGCTGG 1419
 Db 1466 AGAAGGTGCGACCTATTGAGGACATTTGAGCTTTCGAAAGCGTTTGAAGAATTAGGTGCCG 1525
 QY 1420 AGAAATATTTGAACCTGCAATGATTTGTGATGTCAGGTAAGGAAAGGATTCGATATAGATCT 1479
 Db 1526 TGAATATCTATTGAACCTGCAATAACTGTGATGCTCAAGGAAAGGATTCGACATAGACTT 1585
 QY 1480 GATCAAGCTAATATCCGATGCTGTGAACCTCTGTTATCGCAAGCAGCGTGCAGAGT 1539
 Db 1586 AGTAAAGCTCATCTCAGATTCAGTAGGCATACCGGTGATCGCAAGCAGTGGACAGGTAC 1645
 QY 1540 CGCTGATCACTTCTCCGAAAGTCTTTAATGAA-ACCAACGCATCTGCTGCCCTTGCAGCTG 1598
 Db 1646 TCCCGACCACTTTTCCGAGGTGTTTGAAGAGACAAACGATCTGCGCGCTTGTCTGCCG 1705
 QY 1599 GCATTTTCCATCCAAAGAGGTTTCCAAAT--AAGGCTGTTTAAAGAGCACTTGTGAGGA 1656
 Db 1706 GCATTTTCCACCGAAAGAGGTTACCAATCCCAATCTGTGAAAGAGCACTTCAAGAGGA 1765
 QY 1657 AGGGATTGAAGTTAGATTGTAAG 1679
 Db 1766 GGCATAGAAGTCAGGATCTGAG 1788

RESULT 4 SPBC887/c

LOCUS SPBC887 S.pombe chromosome II cosmid c887. 44308 bp DNA linear PLN 01-JUL-1999
 DEFINITION AL033388

ACCESSION 3-oxoacyl-[acyl-carrier-protein]-synthase; 50s ribosomal protein
 VERSION AL033388.1 GI:3850096
 KEYWORDS 114; adenylate cyclase; amidotransferase; calcium-transporting
 leucine Rich repeat; golgi membrane; histidine biosynthesis;
 pif1; protein sorting; pseudouridylylase synthase; response regulator
 mcs4; rna-binding; rpl1; rrm3-pif1 helicase homolog;
 transcriptional activator; WD repeat.

SOURCE ORGANISM

Schizosaccharomyces pombe (fission yeast)
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;
 Schizosaccharomycetes.

REFERENCE AUTHORS

TITLE Direct Submission
 JOURNAL Submitted (04-NOV-1998) European Schizosaccharomycetes genome
 sequencing project, Sanger Centre, The Wellcome Trust Genome
 Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
 and Laboratoire de Recombinaisons Genetiques, 2 avenue du
 Professeur Leon Bernard, Faculte de Medecine, F-35043 Rennes cedex,
 France

COMMENT

Notes:
 Details of yeast sequencing at the Sanger Centre are available on
 the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
 During 1995 to 1996 about 66% of S. pombe chromosome 1 was
 sequenced by the Sanger Centre. The sequencing of the S. pombe
 genome is now being continued with funding from The European
 Commission. Fourteen European sequencing laboratories, including
 the Sanger Centre, are participating in the project.
 Protein coding regions (CDS) have been predicted with the help of
 computer analysis using the Genefinder program in FomBase (an ACSDb
 database) with additional predictions for the branch-acceptor sites
 supplied by the program Sp3splice. CAUTION: It is possible that for
 any individual CDS we may have underestimated or overestimated the

number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC5H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid c887 has no overlaps at present.

FEATURES

source

1. .44308
/organism="Schizosaccharomyces pombe"
/mol_type="genomic DNA"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="II"
/map="IIR"
/clones="cosmid c887"
790. .1326
/gene="SPBC887.01"
790. .1326
/gene="SPBC887.01"
/note="SPBC887.01, len:178, SIMILARITY:Arabidopsis thaliana, O23327, hypothetical 108.0 kd protein, (940 aa), fasta sc ores: opt: 350, E():5.6e-27, (57.1% identity in 84 aa)"
/codon_start=1
/label="SPBC887.01"
/product="hypothetical protein"
/protein_id="CAA21886.1"
/db_xref="GI:3850097"
/db_xref="SPTREMBL:O94286"
/translation="MRAVIFQDEGDQKPNDSKIEVSAEALAKVSYRHHDGLHTF ADGLMKYGFKNRDEVVSVKGLGDRYDNVKKPFPEHLEHDEIRLLIDNGYFQVR SVDDRWRIFFVEKGLDILLPGIYHRTTDDYIHAWRFLFENPKWIALSRDTSSE ELDARKSYLNSIRKSVV"
join(1830. .1959,2071. .3944)
/gene="SPBC887.02"
join(1830. .1959,2071. .3944)
/gene="SPBC887.02"
/note="SPBC887.02, len:667, SIMILARITY:Cavia porcellus, P97 274, volume-regulated outwardly-rectifying chloride channel., (760 aa), fasta scores: opt: 450, E():0, (26.8% identity in 724 aa)"
/codon_start=1
/label="SPBC887.02"
/product="putative chloride channel protein"
/protein_id="CAA21887.1"
/db_xref="GI:3850098"
/db_xref="SPTREMBL:O94287"
/translation="WKDDLLIDWHERYEQKSANRLAGRFPLGLETKYSIISISG IFGLTALLNALSLNSREGCTWNILPDQCTCTTEDEYCEQFFERNHVS FVCLIIYVSVGAFATTLGYVAPARASGPTIAILSGKYPDMNVPSIKTL GSAELVCFVSAGLWKGEGFVHIATNIYLVRIAPSLADSEIFTRQLAAWAS G1AASFAPGVGVIFALEQLASSPFLSTGSIWYEFCLSSAVVALQLIRSWHTDVG YLSVSLDRWYSKOTLPFIPISILCGLAGSLVILYLNKPKASKTGFSKISNFFVFI LSLTSLTDAWYSGELLNPMELFPQVINCSPSSSTVLCETTFWVTAVLFTSAL LGLLTSATGGAAPPTGIIIVPSLAIGACIGRAVGTLLKSPPSLACTSIYGVIGSIAP LSSSTRIVVALVILFETGALNLPMLATLISKWVSISITTSIYDIAMIQPNRI FFPSSSLKFSIPLNFVRSPQLVRPIRSCSIEELERAMHDSQSFFVVLKNDTFE PFGFISRNKYSLELNRPMNSNMOTDNTGLDPLRSASAPVDSFTDLFDYIHPTFTFL NYDTPPVLMKLPKDGATINLALLNKGKLGVLTKIVCFICTI"
1860. .3380
/gene="SPBC887.02"
/note="Pfam match to entry PF00654 voltage_CLC, Voltage gated chloride channels"
1960. .1965
/gene="SPBC887.02"
/note="gtacgt, splice donor sequence"

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

misc_feature

gene

CDS

2055. .2070
/gene="SPBC887.02"
/note="ctgacattttggaag, splice branch and acceptor"
complement(join(4238. .6370,6416. .6526))
/gene="SPBC887.03c"
complement(join(4238. .6370,6416. .6526))
/gene="SPBC887.03c"
/note="SPBC887.03c, len:747, SIMILARITY:Saccharomyces cerevisiae, O07896, chromosome xii reading frame orf ylr002c., (663 aa), fasta scores: opt: 1099, E():1.9e-32, (33.1% identity in 629 aa)"
/codon_start=1
/label="SPBC887.03c"
/product="hypothetical protein"
/protein_id="CAA21888.1"
/db_xref="GI:3850099"
/db_xref="SPTREMBL:O94288"
/translation="MAARKNQSPKPKVASSKLNKIKSKNNSSQSTPRKNATSL DSKTTTKGVALPEIAARELTQEDIEFFNENPSSKYLSSINPDLGKVEKGPDI YDLKKSQFELDTSLRSSDESVLDYSDSEBDQYELAPRVSSWNNSYNRLPIK KGLLONVVADVNGCEFEFLSESEASLEIDSIDKDEKOKSEEQKIAPEIPVKQIK NKALGICAOQQLPEPVENLHIRNIPKFPSPYTIKKLSILLTLLAVPRDIIKGYK IRPLSEEGTKLSKEVAORWEYEQTLKHYAKFLQTLTKLSKFSSTLDETQLSLYQ VAVRCCTKLEIQASHFNSEKLFALAVRQISHTKRPFGDIINSUKNIFEDNLOKT SLKCVTILSRMFKQRYNDVLPDYDLFLSVNINLDMKIDEAWQDDTTNFKRKKLP YLTKKARKYKTKITQEMKADAVITQDKDEKYQSEILKIIFFYFTLQKGLI GNALEGAELSHLNIETFGDLLQVLRELMDVTFLPKDKSGVQATREALLTVSTAF ETASAGVSKLNLDDLGLFVORLYKLIIFPFLNPADNLNLTSLKLPDAPSKPVV NATTEMULLKCFVFFKSNSSRLSKRLAISMQLPHEHSAADLALLKKLL SKYSLSLTSEEIGDGIYNPFTEDPLSNSSTAVLYEPFLKKNHSPAVSQSAKE LKKTSL"
complement(6371. .6381)
/gene="SPBC887.03c"
/note="cttacatttag, splice branch and acceptor"
complement(6410. .6415)
/gene="SPBC887.03c"
/note="gtatga, splice donor sequence"
complement(join(6806. .8727,8788. .8959,9026. .9073))
/gene="SPBC887.04c"
complement(join(6806. .8727,8788. .8959,9026. .9073))
/gene="SPBC887.04c"
/note="SPBC887.04c, len:713, SIMILARITY:Mus musculus, PLAP MOUSE, phospholipase a-2-activating protein, (646 aa), _fast a scores: opt: 1091, E():0, (41.8% identity in 402 aa)"
/codon_start=1
/label="SPBC887.04c"
/product="WD repeat-containing protein"
/protein_id="CAA21889.1"
/db_xref="GI:3850100"
/db_xref="SPTREMBL:O94289"
/translation="MTSYELSRBLGCHKQDVRGVCISNELIGSARDGTYVWEQIN GEWTFPHNHEGFVNCVCPVPAIDKNGRGQKCGILQEVGNTNRSYLYFGHESNIC SASALNETITGSDNSTARVMALGCKYVLKQHSVAVLALGSDIDFITGSAKLI KWNGKLVKLSILAHNDVRSICQIPGGFASCNDDGVIKLTSDTDFLHGHGTFV YSLTYHNQOLIASCEGDIRIRWKGECLOCIITPTTSVWSVSLPGLVHGTGSDG FVPIFTVDRVAPTEVLKNFERVSQFAISSQVGDIKKSLPGLIILSKPGKADQ VNVRVNDEAYQWQSKENWKIKQVDAVGNRQKLPFEGKEYDVFVDVADQQA PLKLPNATENPYQAANRELNLQPLSTYDEVVKTEKTHQHSLESKEPNLESQS SNIKTITTFVSQLLSFNANVPAMCORLRSNNKSNPLPAKSIDSLERALSCKITD TENKELLETCLSLDSLAERFPALDALLRLAINSSDLAPIFLEVFVRVKSVPSS GNPSVNMVLAIRGLSNVNPNTDAEGVKMLDCLTSTVQASSAKDFKIAFATLANN LSTLLQLALENTGIELLSILFSDLPDNEAFYRALMALGTLCTVDPDIALAASQI YHAQSIHVHIAERFSQEMFVDAEKILSLF"
complement(8228. .8341)
/gene="SPBC887.04c"
/note="Pfam match to entry PF00400 G-beta, WD domain, G-bet a repeat"
complement(8345. .8461)
/gene="SPBC887.04c"
/note="Pfam match to entry PF00400 G-beta, WD domain, G-bet a repeat"
complement(8468. .8578)

Kiesau, P., Kirchthath, L., Koetter, P., Korol, S., Liebl, S., Logghe, M., Lohan, A.J.E., Louis, E.J., Li, Z.Y., Maat, M.J., Mallet, L., Mannhaup, G., Messeang, P., Miooga, T., Molemans, P., Mueller, S., Nasr, F., Obermaier, B., Perea, J., Pierard, A., Piravandi, E., Pohl, F.M., Pohl, T.M., Potier, S., Proft, M., Purnelle, B., Ramezani, Rad, M., Rieger, M., Rose, M., Schaaff-Gerstenschlaeger, I., Scherens, B., Schwarzlose, C., Skala, J., Slonimski, P.P., Smits, P.H.M., Souciet, J.L., Steensma, H.Y., Stucka, R., Urestarazu, A., van der Aart, Q.J., van Dyck, L., Vassarotti, A., Vetter, I., Vierendels, F., Vissers, S., Wagner, G., de Weyfousse, P., Wolfe, K.H., Zagulski, M., Zimmermann, F.K., Mewes, H.W. and Klein, K. Complete DNA sequence of yeast chromosome II	QY	910	GGTCAGCTTCTTAAACATTAAGTCTGGTTTCGGGACATTCCTCTAGGCGATCTACCCATGCT	969
EMBO J. 13 (24), 5795-5809 (1994)	Db	1132	AGTAACATTTTGAATATAACTCTTTTAGAGATTGCTCTTTGAAGGATCTCGATGCT	1073
95112788	QY	970	ACAGGCTTTCGACCGGATCTGAAACGTTTGTGCGCATTAACCTGTGGGGGTGGCAT	1029
7813418	Db	1072	AGAGGTTCTGAAACAAAGCGGCAAGACAGCTCTTTGTTCCATTGACAGTCGGTGGGGGAT	1013
2 (bases 1 to 2429)	QY	1030	CAGGGAATTTTACTGATGCAATTAAGAGGTATTATTCTAGTCTAGAAAGTGGCTTCAGAGTA	1089
Aljinovic, G., Pohl, F.M. and Pohl, T.M. Unpublished	Db	1012	CAAGGATATTGTCGATGTTGATGGAACCAAAATACCTGCTTTAGAAAGTGTCAAGTCTATA	953
3 (bases 1837 to 2429)	QY	1090	TTTCAGATCGGCGCGGATTAAGGTTTCGATCGGAAGTGTGATGCTGCTCGAGGA	1149
Aigle, M., Baclet, M.C., Barthe, C., Biteau, N., Crouzet, M. and Doignon, F. Unpublished	Db	952	CTTCAGATCTGGTGTCTGATTAAGTATCGATCGGTACGATGCACTCTATGACGCCGAAA	893
4 (bases 1 to 2429)	QY	1150	ATATATTAAACCGGAGTGAAGACA---GGAAGAGCAGCATAGAGAGATATCTACAGT	1206
Direct Submission	Db	892	ATACTACGAGTTGGGTAACAGAGGAGATGGAAGCTCACCATAAGAGACAATCTCGAAGC	833
Submitted (30-AUG-1994) Data collected by MIPS on behalf of the European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org	QY	1207	ATATGCTAACCGGCGGATTAAGGTTTCGATCGGAAGTGTGATGCTCGCGGAGTTTACTTTGAGAAAAAC	1266
Location/Qualifiers	Db	832	ATACGGTGTCTAGGCGAGTTGTTATTCTGTGCGACCTTAAGAGAGTATATGTAATTCACA	773
1. .2429	QY	1267	CGATGAAGTAGAATTTAAAGCCATCAAGTAAGCCATCCAGGTCCTCAAAACGGTGAGGAATA	1326
/organism="Saccharomyces cerevisiae"	Db	772	AGCAGATACGAGAACAAAGTCTTCGAGACAGAAATATCCGGGCCCAATGAGAGAAATA	713
/mol_type="genomic DNA"	QY	1327	TGCTGCTGATGCTGATGCTGTTTAAATGCTGAGACGAGGAGACCCATCGGAGCTTATGA	1386
/db_xref="taxon:4932"	Db	712	CTGCTGCTACCAATGTACAATCAAGGTGGAAGAGAAATCTAGAGACCTTGCTGTGTGGGA	653
/chromosome="II"	QY	1387	ACTAGCTAAGGCTGTGAGGAATCTCGAGCTCGAGAAATATTATTGAACCTGCAATTGATTG	1446
/complement (352. .2010)	Db	652	ATTAACAAGGCGCATGTGAAGCTCTAGGTCGCGGAGATTTTATTGAACCTGCAATAGCAA	593
/gene="HIS7"	QY	1447	TGATGCTCAAGGAAAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGATCTGTGAA	1506
/complement (352. .2010)	Db	592	GGATGCTCTAATCTGTTTATGATCTGGAATTTAGTAACAATGTTAAAGATCGGTCAA	533
/note="ORF YBR248c"	QY	1507	CATTCTGTTATCGCAAGACGCGGTGAGAGTCTGATGATCTTCCGAAAGTCTTTAA	1566
/codon_start=1	Db	532	GATTCCCGTCATTGTCATCGAGTGGCGCGGTGTACCGCAACATTTCCGAAGAGCCTTCCT	473
/db_xref="GI:536663"	QY	1567	TGAACCAACGATCTGCTCCCTTCAGCTGCGCATTTTCCATCGCAAGAGGTTCCAAT	1626
/db_xref="SWISS-PROT:P33734"	Db	472	AAAGCCCGCGCAGATGCTTGTGCTGAGGTATGTTCCAGAGAGTGAATCACTGT	413
/translation="MPVHVIVDSGNLQSLTNAIHLGYEVQLVSPKDFNLSGTSR LILPGVNGYHFDVNLNFRGEPRIEYIESGPIGMICVLQALFAGSVESPKTGL NYIDFKLSRDFDSEKPYPEIGWNSCIPSENLFFGLDPYKRYFVHSPAILNSEKKQ LENDGWIKAKYGSSEFIAAVNKNIFATQFPEKSGKAGLVNENFLKQSPPIPN VSAFEKLLMNDYSNYGLTRIIACLDRVNDQDLVVTGKGDVREKSDGKGVNLL GKPVLQAKYQOGADSVTLNITSFEDCPKDTPLVLEKQAKTVFVPLTVGGGIK DIVDVGTKIPALEVASLYFRSGADKVSIGTDVYAAEKYELGNRGDGTSPHETISK AYGAQAVISVDPRVYNSQADTKNVFETPYPGNGEKYCWYQCTTKGGRSRLDG VWELTRACALGAILNCIDKGSNSYDLELIEHVDAVKIPVIASSGAGVPEHF EEAFLKTRADACLACAGMFRGEFTVNDVKEVYLLHGLKVRMDEE"	QY	1627	TAAGGCTGTTTAAAGAGACACTTGTGTTGAAGGAAGGATTTGAAGTTAGATTGTAAGCGGAGAA	1686
BASE COUNT	Db	412	TACGATGTAAGGAGTATTACTAGACGCGGATTAAGGTTAGAAATGATGAAGAGTA	353
ORIGIN				
Query Match				
Best Local Similarity				
Matches				
Conservative				
Mismatches				
Gaps				
Indels				
Score				
DB 8				
Length				
2429				
Pred. No.				
5.6e-65				
0				
380				
Indels				
3				
Gaps				
1				
QY		730	TTCTAAACTTCAAGAGAGTAATGCTTGGCTTGTAGTGAGGCAATGATTAATGGGA	789
Db		1312	TTATGGTCTAACACGCGAAGTATTGCTTGTGTTGATGATGCTACTAATGACCAAGTGA	1253
QY		790	TCTTGTGTTACCAAGGAGACCAATATGATGTGAGAGACGCTACAGAGAGATGAGGT	849
Db		1252	TTTGTGTTACTAAGGTGATCAATACGATGATGCGTGAAGAAAGTGTGTAAGGTGT	1193
QY		850	CAGAAACCTTGGCAAGCTGTTGAACTTGTCTGGGAGTATTATTATTAGACGGTGTCTGATGA	909
Db		1192	TAGAAACCTTGGTAAAGCTGTTTCAAGTTGGCACAGAAATATTACCAACAGGGTGGGATGA	1133
LOCUS				
SCHIS7				
DEFINITION				
S.cerevisiae HIS7 gene.				
ACCESSION				
X69815				
VERSION				
X69815.1				
GI:395154				
KEYWORDS				
glutamine amid transferase.				
Saccharomyces cerevisiae (baker's yeast)				
SOURCE				
ORGANISM				
Saccharomyces cerevisiae				
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
REFERENCE				
1				
AUTHORS				
Kuenzler, M., Balmelli, T., Egli, C.M., Paravicini, G. and Braus, G.H.				
TITLE				
Cloning, primary structure, and regulation of the HIS7 gene encoding a bifunctional glutamine amidotransferase: cyclase from Saccharomyces cerevisiae				
JOURNAL				
J. Bacteriol.				
175 (17), 5548-5558 (1993)				
MEDLINE				
93374850				
PUBMED				
8366040				
REFERENCE				
2 (bases 1 to 2440)				

/db_xref="GI:28950247"
/translation="MSDKPSDQTPNQRRSSRSRSPAGPVRDRDRDHWKNTGGA
QKAFSLNDOQRIIRLELAQOCTQVGRREPDLERLNQDPSPSPHCFNMARQDSFLK
PNAYLAKTIDDDOPAAEVAKAPSVSDVGVGNKCKGRKGTVAQCILIPDOPGLVYG
DOEPALTDVDDOPAAEVAKAPSVSDVGVGNKCKGRKGTVAQCILIPDOPGLVYG
CPICNTDHLVDNCPKPKPAPGTQLDPKVIQBYIERLDVCFVKANRPPFATNGM
TWGLFTGYRHAHINTFELNKSMDTGRHYPTPEYAMRWVDSEVERDMKKFDP
TTTCRNVCSSHSDPVPYKYESISETIDAYLDGKWDVATGMIPTTNVDRLTRLWY
LEAARLVFSTQVFLKEPECAEAPAPVPKPSFPPTIVSNTSNLVQVYVWNEETQNF
VPDTSFSLFSVTTQKAPNLLRCPDEADYAJEYAAWFKHQRPNQPGVGDLEDOA
LRAVIRAEGLRNRERVAKLGRQAQKRRNAREADREMAFEAQNASTIDVHKREH
D"
complement (18190..19685)
/gene="103E1.040"
/number=1
complement (19686..19790)
/gene="103E1.040"
/number=1
complement (19791..19998)
/gene="103E1.040"
/number=2
21036..21715
/gene="103E1.050"
complement (join(21036..21062,21176..21715))
/gene="103E1.050"
complement (join(21036..21062,21176..21715))
/gene="103E1.050"
/notes="similarity to hypothetical 24.1K protein,
Neurospora crassa, PIR:JH0145"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD71114.1"
/db_xref="GI:28950248"

Query Match 15.58; Score 303.2; DB 8; Length 87435;
Best Local Similarity 58.24; Pred. No. 1.4e-57;
Matches 552; Conservative 0; Mismatches 393; Indels 3; Gaps 1;

QY 747 GAGTAATTGCTGCTTGATGTGAGGCAATGATATGGGATCTTGTGTAACCAAGG 806
DB 32317 GTGTCATTGCTGCTGGATGTGGCAACCAACGACGAGCGGACTTGGTGTGTCACCAAGG 32258
QY 807 GAGACCAATATGATGTGAGAAACGTCACGAGAGATGAGGTGAGAAACCTTGGCAAGC 866
DB 32257 GTGACCAATATGATGTCCGGAAGGCTGACGACGCAACGTCGCAACCTCGGCAAGC 32198
QY 867 CTGTTGAACCTTGGCGCAGTATATTTAGACGGTGTGATGAGGTGAGGTCTTCTTAAACA 926
DB 32197 CCGTCGAGATGGCGCGCAAGTACTACGAAACGCGCGGACGAGTACCTTCTCAACA 32138
QY 927 TTAAGTGTTCGGGACTTCCCTCTAGCGGATCTACCCATGCTACAGGTCTTGCACGCG 986
DB 32137 TCACCTCTTCCGGGACTGCCCGTTCGCCACCTGCCCATGCTCGAGATCTTCCGCTGA 32078
QY 987 CATCTGAAACGTTTTTGTGCCATTAACCTGCGGGGTGGCATCAGGATTTTACTGATG 1046
DB 32077 CCTCCAAGACGCTTGTCTTCCCTGACCTGCGCGCGCATCCGCGACACCGTCGACA 32018
QY 1047 CAATGGAAGGTATATTTAGTCTAGAAAGTGGCTTACAGATATTTACAGATCGGGCGCG 1106
DB 32017 CGGACGCAACCAAGGTCTCCGCGCTCGAAATCGCAACATGATCTTCCAGTCAGGCGCG 31958
QY 1107 ATAAGTTTCGATCGGAAGTGTGAGTGTACCTGCTGAGGAATATATTAACCGGAG 1166
DB 31957 ACAAGGTGTCCATTTGGCTCCGACGCGCTATCGCTGCCGAGGAGTACTATGCTCCGCGCA 31898
QY 1167 TGAAG---ACAGAAAGACGACATAGACGAGATATCTACAGTATATGTTAAACGAGCAG 1223
DB 31897 AGACGCTCTTTGGCAACACCGGCATTTGAGCAGATCAGCAAGCTACGGAACACGAGCTG 31838
QY 1224 TGGTTGTAAGCATGATCTCCGCGAGTTTACTTGTGAGAAACCCGATGAAGTGAATTTA 1283
DB 31837 TCGTGTCTCGGTTCGATCCCAAGCGGTCTACGTCGCCAAGCCGATGTCACGACGACACC 31778

QY 1284 AAGCCATCAAGTAAGCCCATCCAGGTCCTCAACCGGTGAGGAATATGCTGTATCAGTGCA 1343
DB 31777 ACACCGTCAAGACGTCGTACCCCGGCCCAAGSGCGAGGAATACTGCTGTATCGCTGCA 31718
QY 1344 CTGTTAATGTTGACAGAGAGGAGACCCCATCGAGCTTATGAACTAGCTAAGCGCTGTTG 1403
DB 31717 CGATCAAGGCGCGGCGGAGACGAGGAGATCTGGATGTGGTTGAACCTCAGCAGCGCGTGG 31658
QY 1404 AGGAACCTTGAGCTGAGAGAAATATTATGAACCTGCAATGATCTGTATGGTCAAGGAAAG 1463
DB 31657 AAGCAATGGCTCGCGGAGATCTGTTGAATCTGCAATGACAGGACGCGACCAACAGCG 31598
QY 1464 GATTTCGATATGATCTGATCAAGCTAATATCCGATGCTGTGAACATTTCTGTTATCGCAA 1523
DB 31597 GGTTCGATCTGGAGTTGATCAGGACGTCAGGCGCGCGTGGAGATACCGGTTATTGGCCA 31538
QY 1524 GCAGCGGTGAGAGTGGCTGATCACTTCTCCGAAGTCTTTATGAACCAACGATCTG 1583
DB 31537 GCTCGGGGCGGCAACCCGCGGTCACTTTGAGGAGGTGTTTAGGAGACGACGACGACG 31478
QY 1584 CTGCCCTTGACGCTGCGCATTTTCCATCGCAAGAGGTTCCAATTAAGGCTGTTAAAGAGC 1643
DB 31477 CCGCTGAGGAGCAGCATGTTCCACCGTGGGAGTATACGATACAGCAGGTCAAGGAGG 31418
QY 1644 ACTTGTGAAAGGAGGATTTGAAGTTAGATTGTAAGCGCGAGATCACT 1691
DB 31417 AGCTGAAGCGGAGGCTGTTGCTCAGACAGTTTTCAGGAGATCTCT 31370

RESULT 8
AF159463
LOCUS
DEFINITION
EMERICELLA nidulans glutamine amidotransferase:cyclase (hishF)
gene, complete cds.
ACCESSION
AF159463
VERSION
AF159463.1
KEYWORDS
GI:8886089
SOURCE
EMERICELLA nidulans (anamorph: Aspergillus nidulans)
ORGANISM
EMERICELLA nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
REFERENCE
1 (bases 1 to 4351)
AUTHORS
Valerius, O., Draht, O., Hoffmann, B., Kuebler, E. and Braus, G. H.
TITLE
Isolation and characterization of the hisF gene of Aspergillus
nidulans
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 4351)
AUTHORS
Valerius, O., Draht, O., Hoffmann, B., Kuebler, E. and Braus, G. H.
TITLE
Direct Submission
JOURNAL
Submitted (16-JUN-1999) Molecular Microbiology, Georg August
University, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
Location/Qualifiers
1..4351
source
/organism="Emericella nidulans"
/mol_type="genomic DNA"
/strain="A234"
/db_xref="taxon:162425"
/note="anamorph: Aspergillus nidulans"
<519..52329
/gene="hishF"
join(<519..641,728..2176,2240..2329)
/gene="hishF"
/product="glutamine amidotransferase:cyclase"
join(519..641,728..2176,2240..2329)
/gene="hishF"
/codon_start=1
/product="glutamine amidotransferase:cyclase"
/protein_id="AAF80376.1"
/db_xref="GI:8886090"

ILEKDWLVARRVYBEKEFIGAIARDNIFATOPHPEKSGQGRPHPSRFLGRSSAPFC
HURDSILTEKNGELTRIIIACTLDVTRNDVGLVTVTKGDVDRKEDGADAGGVORNLG
KPYDMAKYEEQADEVTEPLNITSFNCPLADLPMLLEILRRTSEVFPVLTIGGGIRD
TVTDGTHIPALDVASMIYFSGADKSVISGSDAVAAEDYIYAGKVLGSKTAIETISKA
YGNQAVVSVDPKRVVSVQPEDTKHRTIETKFPNAAQGNFCWPTQCTIKGKRETRDLUDV
COLVQAVEMAGAEIILLNCIDKDGNSGSLTDQPEFSKRAVKIPVITASSGAGMPKHF
EVPDOTTDAALGAGHFRGEYTVGVKQVLEDRGLVRRFEDV"

BASE COUNT 1138 a 975 c 1097 g 1141 t
ORIGIN

Query Match 13.6%; Score 266.6; DB 8; Length 4351;
Best Local Similarity 55.7%; Pred. No. 2.4e-49;
Matches 581; Conservative 0; Mismatches 444; Indels 18; Gaps 3;

QY	578	GAGCATTTTATTCATCATTCAGAGGGAATGTTTCATCGACTCCAAATTCATCCCGAG	637
DB	1133	GAGAAATTCATCGCGCGCATAGCAGAGATAACATCTTTCGACACAATTCATCCCGAG	1192
QY	638	AAGAGTCGAGGTGTGGACTTTCATATTGAGAAGATTTT-----GAATGCTGATTC	691
DB	1193	AAGAGCGGCACAGGCAGCGCTAGCCACCTTCGCGCTTCTTGACGCGAGCTCAGCTCCA	1252
QY	692	TTTAAACAAAAGACAGAACCAATGAATGGAAGGCTTCTAACTGTCAAAAGAGATGA	751
DB	1253	TTCTGTACATTTAGAGACTCGATTTTGACAGGAGAGAAAAACGCTCTTACCCTAGGATC	1312
QY	752	ATTGCTTCCTTGATGTGAGGCAATGATTAATGGGATCTTGTGTAAACCAAGGAGAC	811
DB	1313	ATCGCTGTCTTGATGTTCGTACGAATGATGTGCGGATCTCGTGTGATAAGGCGCAT	1372
QY	812	CAATATCATGTGAGAGAACGTACAGAGAGAAATG-----AGTTCAGAAACCTTGGC	862
DB	1373	CAATATGATGTTCGAGAGAGATGTGCGGATGCTGGAGGCAAGTGAGAACCTGGGA	1432
QY	863	AAGCCTGTGAACTTGTGCGGAGTATTAATTTAGACGGTCTGATGAGGTGACGTTCTTA	922
DB	1433	AAGCGGTGTGATGTGCTAAGAAATAATACGACAGGGGCGAGATGAGGTGACGTTTAA	1492
QY	923	AACATTAAGTGTTCGCGGACTTCCCTCTAGGCGATCTACCCATGCTACAGGTCTTCAA	982
DB	1493	AACATCACCTCTTTTCAGAAACTGTGCGGTAGCCGACCTCCCTATGCTCGAGATTCCT	1552
QY	983	CGGCATCTGAAACAGTTTTGTGCCATTAACCTGCGGGGTGGCATCAGGGAATTTTACT	1042
DB	1553	AGAAGCTGGAGACCGTCTTCGTAACCTTGACTATTTGTTGGCGCATAGGACACTGTG	1612
QY	1043	GATGCAAAATGGAAGGTATTAATTTAGTCTAGAGTGGCTTCAGAGTATTTCAATCGGG	1102
DB	1613	GATACAGACGTACTCATATCCAGCTCTAGACGTGGCATCGATGTAATTTCAAACTGGG	1672
QY	1103	GCGCATTAAGGTTCGATCGGAAGTATGTCAGTTTACATGCTGAGGAATATTAATAACC	1162
DB	1673	GCTGACAAAGTCAGCATTTGGTTGCGATGCCGTGTTGCTCGGAAGATTAATACGACGT	1732
QY	1163	GG----AGTGAAGACAGAAAGACAGCATAGACAGATATCTACAGTATATGTTAAACC	1219
DB	1733	GGCAAGTGTCTGTCTGGCAAACTGCGCAATGAACTATTTCTAGGCGTATGGAACCAAG	1792
QY	1220	GCAGTGGTTGTAAGCATTTGATCTCGCGAGTTTATCTTGAAAAACCCGATGAAGTAGAA	1279
DB	1793	GCTGTGTTGTAAGGTTGACCCGAGCGGTTTTATGTTCAGCCAAACAGAACACAGAA	1852
QY	1280	TTTAAAGCATCAAAAGTAAGCCATCCAGGTCCAAACGGTGAAGGAATATGCTGTATCAG	1339
DB	1853	CACCGTACGATAGAAACGAAATTTCTTAAACCGCCCGGCAAAATTTCTGTGTGATCCAG	1912
QY	1340	TGCACTGTATATGTGACGAGAGAGGAGACCCATCGGAGCTTATGAATAGCTTAAGGCT	1399
DB	1913	TGTACTATAAAGGTGGCAGAGAGACCCAGACTTAGATGCTGCCAGCTGTGTCAGGCC	1972
QY	1400	GTTGAGGAACCTTGGAGCTGGAGAAATATTAATTTGAACCTGCAATGATGTGATGGTCAAGGA	1459
DB	1973	GTCGAGGCAATGGGTGCTGGGAGATTCTGTGTAATTTGCAATTAAGACCGGAGCAAC	2032

QY	1460	AAAGGATTCGATATATAGATCTGATCAAGCTAAATATATCGATGCTGTGAACATTCCTGTTATC	1519
DB	2033	AGTGGTTTCGATCTTGACTGATCAACACCGCTCAAGCGGTGAGTAAATAATACCCGTGATT	2092
QY	1520	GCAGCAGCGGTGACAGGAGTCCGTGATCACTTCTCGAAGTCTTTTAAATGAAACCAACGCA	1579
DB	2093	GCCTCTAGCGAGCTGGTATGCGGAAGCATTTTGGAGGAAGTTTGTGATCAAAACGACGACA	2152
QY	1580	TCGTGCTGCCCTTGACGCTGGCAT	1602
DB	2153	GATGCTGCTCTGGTCTGGGAT	2175

RESULT 9

AC138524.1/c

WPCOMMENT

Sequence split into 6 fragments LOCUS AC138524 Accession AC138524

Fragment Name	Begin	End
AC138524_0	1	110000
AC138524_1	100001	210000
AC138524_2	200001	310000
AC138524_3	300001	410000
AC138524_4	400001	510000
AC138524_5	500001	550277

Continuation (2 of 6) of AC138524 from base 100001 (AC138524 Homo sapiens chromosome 5

Query Match 12.0%; Score 234.8; DB 2; Length 110000;

Best Local Similarity 60.5%; Pred. No. 3.6e-42;
Matches 422; Conservative 0; Mismatches 272; Indels 4; Gaps 2;

QY	730	TTCTAAACTTGCAGAGAGATTAATTCCTTCCTTGATGTGAGGGCAAAATGATATGGGA	789
DB	21179	TTATGCTCTAACACACGAGAAATTAATTCCTTGTCTTGATGACGTACTAATGACCAAGTGA	21120
QY	790	TCTTGTGTAAACCAAGGAGACCAATATGATGTGAGAGAACGTACAGAGAGATGAGGT	849
DB	21119	TTTGGTGGTTACTTAAAGGTGATCAATACGATGTACGTGAAAAAAGTGATGTTAAAGTGT	21060
QY	850	CAGAAACCTTGGCAACCTGTTGAACTGCTGGCGAGTATTATTAGACGGTCTGATGA	909
DB	21059	TAGAAACCTTGGTAACCTGTTGACGTGGCAAGAAATATTACCAACAGGGTGGGATGA	21000
QY	910	GGTCAGCTTCTTAAACATTAATCTGTTTCCGGGACTTCCCTCTAGGGCATCTACCCATGCT	969
DB	20999	AGTAACATTTTGAATATAACTTCTTTTAGAGATTGTCCTTTGAAGGATACTCCGATGCT	20940
QY	970	ACAGGCTTGCACACGGCGCATCTGAAAAAGTCTTTTGTGCCATTAACCTGTCGGGGTGGCAT	1029
DB	20939	AGAGGTTCTGAAACAAAGCCGCAAGACAGTCTTTTGTTCATTTGACAGTCCGTGGGGGGAT	20880
QY	1030	CAGGGATTTTACTGATGCAAAATCGAAGGTATTATTCTAGTCTAGAGTGGCTTCAGAGTA	1089
DB	20879	CAAGGATATTGTCGATGTTGATGGAACCAAAATACCTGCTTTAGAGTCCGAAGTCTATA	20820
QY	1090	TTTTCAGATCGGGCGCGATAAGGTTTCGATCGGAAGTATGTCAGTATTACACTGCTGAGGA	1149
DB	20819	CTTCAGATCTGGTGTCTGATAAAGTATCGATCGGTACGATGCGATCTATCAGCCGAAAA	20760
QY	1150	ATATATTAAACCGGAGTGAAGACA---GGAAGAGCAGCATAGAGCAGATATCTACAGT	1206
DB	20759	ATACTATGAGTGGGTAAACAGAGAGATGGAAGCTCAACCAATAGAGACAATCTGGAAGC	20700
QY	1207	ATATGTTAAACCGGAGTGGTTCGTAAGCATTTGATCTCGCGGAGTTTACTTTGAGAAAAACC	1266
DB	20699	ATACGGTGTCTCAGCAGTGTGTTATTTCTGCGACCTTAAGAGAGATATATGTCATTCACA	20640
QY	1267	CGATGAAGTGAATTTTAAAGCCCATCAAGTAAGCCATCCAGGTCCAAACCGGTGAGGAATA	1326
DB	20639	AGCAGATACCAAGAACAAACGCTCTTCGAGACAGAAATATCCGGGCCCAATCGAGAGAAATA	20580
QY	1327	TGCTGTGT-ATCAGTGCACGTGTTAATGCTGACGAGAGAGGAGGCCCATCGGAGCTTATG	1395

AUTHORS Day, T. W. and Davison, V. J.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUL-2000) Medicinal Chemistry and Molecular Pharmacology, Purdue University, 1333 Robert Heine Pharmacy Building, West Lafayette, IN 47907-1333, USA

FEATURES
 source 1. .2306
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"
 <161. .>2044
 /gene="HIS7"
 <161. .>2044
 /gene="HIS7"
 /product="imidazole glycerol phosphate synthase"
 .161. .2044

gene
 mRNA
 CDS

BASE COUNT 844 a 327 c 376 g 759 t

ORIGIN

Query Match 10.1%; Score 198.2; DB 8; Length 2306;
 Best Local Similarity 54.9; Pred. No. 6e-34;
 Matches 535; Conservative 0; Mismatches 383; Indels 57; Gaps 5;

Qy 741 CAAGAGAGTAATGCTTGGCTCATGTGAGGCAAAATGATTAATGAGGATCTTGTGTGTA 800
 Db 1029 CTAGAAGATTAATGAGTCTGATGTTAGTAATGATGATGATGATGATGATGATG 1088

Qy 801 CCAAGGAGGACCAATATGATGTGAGAACGTACA----- 835
 Db 1089 CTAAGGTGATCAATATAATGTTCCGAGAACAAATTCATCATCATCAATCAACGAATCTG 1148

Qy 836 -----GAAGAGATGAGTCAAGAACCTTGGCAAGCCTGTTGAACCTGCTGGGCAGT 887
 Db 1149 CCGAGGAGGATCTTCAAGGTCGTAAATTTAGGTAAACAGGTTGAATAGCAACAAAT 1208

Qy 888 ATTATTAGACGGTGTGATGAGTCAAGTCTTAAACATTTACTGTTTCCGGGACTTC 947
 Db 1209 ATTATAATCAAGGAGCTGATGAATATACATTTTAAATATACATCTTTCGTAATCTC 1268

Qy 948 CTCTAGGCGATCTACCATGCTACAGGTCTTGCAACCGCATCTGAAACGTTTTTCTGTC 1007
 Db 1269 CTTTAAAGATTTACCCATGTTTCAAGTATTGAAAGAGCTGCTGAAATATATTCTTC 1328

Qy 1008 CATTAAGTCTCGGGGTGGCATCAGGATTTTACTGAT---GCAATGGAAGTATTATT 1064
 Db 1329 CATTAAGTCTGTTGGGAGGATTAAGAGATTTATATGATCCAGAACTGGGAAACCTGTC 1388

Qy 1065 CTAGCTAGAGTGGCTTCAAGATTTTCAGATCGGCGCGATGAAGTTTCGATCGGAA 1124
 Db 1389 CTGCTGTTGATGGCTCATTTATTTCCAAATCAGGTGCTGATAAATAGTATTGAT 1448

Qy 1125 GTGATCAGGATTTACACTGCTGAGGAATATATT---AAACCGGAGTGAAGACAGGAAGA 1181
 Db 1449 CTGATCGCGTTACCATTGCCGAAATATTATTATGCAATTAATGCGGAAATATACCGGTCTTT 1508

Qy 1182 GCAGCATAGACAGCATATCTACAGTATATGTTAAACAGGACAGTGGTTGTAAGCATTCATC 1241
 Db 1509 CATCAATTAAGAACCATTTCTTCCAAATTTGGTAATCAACAGCATGTTTATATCATGTTGATC 1568

Qy 1242 CTCGCCAGTCTTACTTTGAGAAAC-----CCGATGAAGTAGAATTTAAAG 1286
 Db 1569 TTAAGAGAAATATATATCAAGATCCAACTGAACTACTACTATGCAACAAATAAAAA 1628

Qy 1287 CATCAAGTAAGCCATCCAGGTCCAAACGGTGAGGAATATGCCTGGTATCATGTCAGTCACATG 1346
 Db 1629 TCACCTGATCTTTCACAAATATGGTCCAAATGGTGGAACAATATTGTTTATTATCAAGTAACCT 1688

Qy 1347 TTAATCGTGCAGAGAGGAGACCCCATCGAGCTTATGAACCTAGTAAAGGCTGTTTGAGG 1406
 Db 1689 CTCAGGTGGAAGGAAATTCATGAATAGGTGCATTAATATGTTTGTAGCTTGTGAAA 1748

Qy 1407 AACTTGGAGCTGGAGAAATATTATTGAACCTGATGATGATGATGATGATGATGATGATGATGATG 1466
 Db 1749 AATTGGGAGCTGGAGAAATTTTATTAAATTTCAATCATGATGATGATGATGATGATGATGATGATG 1808

Qy 1467 TCGATATAGATCTGATCAAGCTAAATATCCGATCCTGTCGTAACATTTCTCTGTTATCGCAAGCA 1526
 Db 1809 TTAATTTTACAATTTATTACAACTAAATTAATCTCAAGTTTCAATCCCCGGGATTCGTAGTA 1868

Qy 1527 GCGGTGCAGAGTCGCTGATCACTTCTCGAAAGTCTTT---AATGAAACCAACGCATCTG 1583
 Db 1869 GTGGTCTGGTAATCTTCAACATTTTCAACAGATTTTGAATGATGATGATGATGATGATGATGATG 1928

Qy 1584 CTGCCCCTTCAGCTGCATCTTTCATCGCAAGAGGTTTCCAAATGAAGGCTGTTTAAAGAGC 1643
 Db 1929 CTGCTTTAGTGTGCTGTTTTCATCGTGGTGAATATACCGTTAATCAAGTAAGAAAT 1988

Qy 1644 ACTTGTGAAGGAG 1658
 Db 1989 ATTACAACAGGAAG 2003

RESULT 12

AB016783
 LOCUS Arabidopsis thaliana 4581 bp DNA linear PLN 06-FEB-1999
 DEFINITION partial cds.

ACCESSION AB016783
 VERSION 1 GI:3413533
 KEYWORDS glutamine amidotransferase/cyclase.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsia.

REFERENCE 1 (sites)
 Fujimori, K. and Ohta, D.
 An Arabidopsis cDNA encoding a bifunctional glutamine
 amidotransferase/cyclase suppresses the histidine auxotrophy of a
 Saccharomyces cerevisiae his7 mutant
 FEBS Lett. 428 (3), 229-234 (1998)
 98316711

JOURNAL MEDLINE
 PUBMED 9654139
 REFERENCE 2 (bases 1 to 4581)
 Fujimori, K.
 TITLE Direct Submission
 JOURNAL Submitted (05-AUG-1998) Ko Fujimori, Osaka Bioscience Institute,
 Department of Molecular Behavioral Biology; 6-2-4 Furuedai, Suita,
 Osaka 565-0874, Japan (E-mail: fujimori@obi.or.jp,
 Tel: +81-6-872-4851, Fax: +81-6-872-2841)

FEATURES
 source 1. .4581
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 join(2069. .2255,2348. .2430,2524. .2618,2711. .2814,
 2928. .3212,3313. .3407,3521. .3619,3800. .3862,3970. .4032,

```
4125. .4207,4338. .4446,4547. .>4581)
/codon_start=1
/product="glutamine amidotransferase/cyclase"
/protein_id="BAJ32287.1"
/db_xref="GI:3413534"
/translation="METAAPFSSIVSRNFFSSSSIRASSPASLFLSKSIGNNVR
KFSPSLRSVRASTSDSVTLIDYAGNVRISRNALRHIGSFIDKQVTFDILNADR
LIFPGVPFAPADVLNRTGMALCKYIENDRPFLICIGLQLLDSFQNGPVRGL
GVIPGIVGRFDAGIRVPHIGNALQVGDSDILDDVGNRHVVFVHSVRAIPSDNKG
DWISSTCYGESFISIRGNVHAVOPHPKSGEGLSVLRFLPKLPATOKPMGSK
ASKLAKVIACLDVRTNDGDLAVTQGDQDVREQSNENEVRLGKPVDLAGQYXDG
ADSEFLNITGFRDFPLGDMIQVLQTSKNVFLVTVGGGRDFTDASGRYISLE
VAIEYFRSGADNIGSDAVFAAEFFIKSVGKTKGSKLSLEIS"
BASE COUNT      1349 a   782 c   816 g   1634 t
ORIGIN
      1
      2
      3
      4
      5
      6
      7
      8
      9
     10
     11
     12
     13
     14
     15
     16
     17
     18
     19
     20
     21
     22
     23
     24
     25
     26
     27
     28
     29
     30
     31
     32
     33
     34
     35
     36
     37
     38
     39
     40
     41
     42
     43
     44
     45
     46
     47
     48
     49
     50
     51
     52
     53
     54
     55
     56
     57
     58
     59
     60
     61
     62
     63
     64
     65
     66
     67
     68
     69
     70
     71
     72
     73
     74
     75
     76
     77
     78
     79
     80
     81
     82
     83
     84
     85
     86
     87
     88
     89
     90
     91
     92
     93
     94
     95
     96
     97
     98
     99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
    701
    702
    703
    704
    705
    706
    707
    708
    709
    710
    711
    712
    713
    714
    715
    716
    717
    718
    719
    720
    721
    722
    723
    724
    725
    726
    727
    728
    729
    730
    731
    732
    733
    734
    735
    736
    737
    738
    739
    740
    741
    742
    743
    744
    745
    746
    747
    748
    749
    750
    751
    752
    753
    754
    755
    756
    757
    758
    759
    760
    761
    762
    763
    764
    765
    766
    767
    768
    769
    770
    771
    772
    773
    774
    775
    776
    777
    778
    779
    780
    781
    782
    783
    784
    785
    786
    787
    788
    789
    790
    791
    792
    793
    794
    795
    796
    797
    798
    799
    800
    801
    802
    803
    804
    805
    806
    807
    808
    809
    810
    811
    812
    813
    814
    815
    816
    817
    818
    819
    820
    821
    822
    823
    824
    825
    826
    827
    828
    829
    830
    831
    832
    833
    834
    835
    836
    837
    838
    839
    840
    841
    842
    843
    844
    845
    846
    847
    848
    849
    850
    851
    852
    853
    854
    855
    856
    857
    858
    859
    860
    861
    862
    863
    864
    865
    866
    867
    868
    869
    870
    871
    872
    873
    874
    875
    876
    877
    878
    879
    880
    881
    882
    883
    884
    885
    886
    887
    888
    889
    890
    891
    892
    893
    894
    895
    896
    897
    898
    899
    900
    901
    902
    903
    904
    905
    906
    907
    908
    909
    910
    911
    912
    913
    914
    915
    916
    917
    918
    919
    920
    921
    922
    923
    924
    925
    926
    927
    928
    929
    930
    931
    932
    933
    934
    935
    936
    937
    938
    939
    940
    941
    942
    943
    944
    945
    946
    947
    948
    949
    950
    951
    952
    953
    954
    955
    956
    957
    958
    959
    960
    961
    962
    963
    964
    965
    966
    967
    968
    969
    970
    971
    972
    973
    974
    975
    976
    977
    978
    979
    980
    981
    982
    983
    984
    985
    986
    987
    988
    989
    990
    991
    992
    993
    994
    995
    996
    997
    998
    999
    1000
    1001
    1002
    1003
    1004
    1005
    1006
    1007
    1008
    1009
    1010
    1011
    1012
    1013
    1014
    1015
    1016
    1017
    1018
    1019
    1020
    1021
    1022
    1023
    1024
    1025
    1026
    1027
    1028
    1029
    1030
    1031
    1032
    1033
    1034
    1035
    1036
    1037
    1038
    1039
    1040
    1041
    1042
    1043
    1044
    1045
    1046
    1047
    1048
    1049
    1050
    1051
    1052
    1053
    1054
    1055
    1056
    1057
    1058
    1059
    1060
    1061
    1062
    1063
    1064
    1065
    1066
    1067
    1068
    1069
    1070
    1071
    1072
    1073
    1074
    1075
    1076
    1077
    1078
    1079
    1080
    1081
    1082
    1083
    1084
    1085
    1086
    1087
    1088
    1089
    1090
    1091
    1092
    1093
    1094
    1095
    1096
    1097
    1098
    1099
    1100
    1101
    1102
    1103
    1104
    1105
    1106
    1107
    1108
    1109
    1110
    1111
    1112
    1113
    1114
    1115
    1116
    1117
    1118
    1119
    1120
    1121
    1122
    1123
    1124
    1125
    1126
    1127
    1128
    1129
    1130
    1131
    1132
    1133
    1134
    1135
    1136
    1137
    1138
    1139
    1140
    1141
    1142
    1143
    1144
    1145
    1146
    1147
    1148
    1149
    1150
    1151
    1152
    1153
    1154
    1155
    1156
    1157
    1158
    1159
    1160
    1161
    1162
    1163
    1164
    1165
    1166
    1167
    1168
    1169
    1170
    1171
    1172
    1173
    1174
    1175
    1176
    1177
    1178
    1179
    1180
    1181
    1182
    1183
    1184
    1185
    1186
    1187
    1188
    1189
    1190
    1191
    1192
    1193
    1194
    1195
    1196
    1197
    1198
    1199
    1200
    1201
    1202
    1203
    1204
    1205
    1206
    1207
    1208
    1209
    1210
    1211
    1212
    1213
    1214
    1215
    1216
    1217
    1218
    1219
    1220
    1221
    1222
    1223
    1224
    1225
    1226
    1227
    1228
    1229
    1230
    1231
    1232
    1233
    1234
    1235
    1236
    1237
    1238
    1239
    1240
    1241
    1242
    1243
    1244
    1245
    1246
    1247
    1248
    1249
    1250
    1251
    1252
    1253
    1254
    1255
    1256
    1257
    1258
    1259
    1260
    1261
    1262
    1263
    1264
    1265
    1266
    1267
    1268
    1269
    1270
    1271
    1272
    1273
    1274
    1275
    1276
    1277
    1278
    1279
    1280
    1281
    1282
    1283
    1284
    1285
    1286
    1287
    1288
    1289
    1290
    1291
    1292
    1293
    1294
    1295
    1296
    1297
    1298
    1299
    1300
    1301
    1302
    1303
    1304
    1305
    1306
    1307
    1308
    1309
    1310
    1311
    1312
    1313
    1314
    1315
    1316
    1317
    1318
    1319
    1320
    1321
    1322
    1323
    1324
    1325
    1326
    1327
    1328
    1329
    1330
    1331
    1332
    1333
    1334
    1335
    1336
    1337
    1338
    1339
    1340
    1341
    1342
    1343
    1344
    1345
    1346
    1347
    1348
    1349
    1350
    1351
    1352
    1353
    1354
    1355
    1356
    1357
    1358
    1359
    1360
    1361
    1362
    1363
    1364
    1365
    1366
    1367
    1368
    1369
    1370
    1371
    1372
    1373
    1374
    1375
    1376
    1377
    1378
    1379
    1380
    1381
    1382
    1383
    1384
    1385
    1386
    1387
    1388
    1389
    1390
    1391
    1392
    1393
    1394
    1395
    1396
    1397
    1398
    1399
    1400
    1401
    1402
    1403
    1404
    1405
    1406
    1407
    1408
    1409
    1410
    1411
    1412
    1413
    1414
    1415
    1416
    1417
    1418
    1419
    1420
    1421
    1422
    1423
    1424
    1425
    1426
    1427
    1428
    1429
    1430
    1431
    1432
    1433
    1434
    1435
    1436
    1437
    1438
    1439
    1440
    1441
    1442
    1443
    1444
    1445
    1446
    1447
    1448
    1449
    1450
    1451
    1452
    1453
    1454
    1455
    1456
    1457
    1458
    1459
    1460
    1461
    1462
    1463
    1464
    1465
    1466
    1467
    1468
    1469
    1470
    1471
    1472
    1473
    1474
    1475
    1476
    1477
    1478
    1479
    1480
    1481
    1482
    1483
    1484
    1485
    1486
    1487
    1488
    1489
    1490
    1491
    1492
    1493
    1494
    1495
    1496
    1497
    1498
    1499
    1500
    1501
    1502
    1503
    1504
    1505
    1506
    1507
    1508
    1509
    1510
    1511
    1512
    1513
    1514
    1515
    1516
    1517
    1518
    1519
    1520
    1521
    1522
    1523
    1524
    1525
    1526
    1527
    1528
    1529
    1530
    1531
    1532
    1533
    1534
    1535
    1536
    1537
    1538
    1539
    1540
    1541
    1542
    1543
    1544
    1545
    1546
    1547
    1548
    1549
    1550
    1551
    1552
    1553
    1554
    1555
    1556
    1557
    1558
    1559
    1560
    1561
    1562
    1563
    1564
    1565
    1566
    1567
    1568
    1569
    1570
    1571
    1572
    1573
    1574
    1575
    1576
    1577
    1578
    1579
    1580
    1581
    1582
    1583
    1584
    1585
    1586
    1587
    1588
    1589
    1590
    1591
    1592
    1593
    1594
    1595
    1596
    1597
    1598
    1599
    1600
    1601
    1602
    1603
    1604
    1605
    1606
    1607
    1608
    1609
    1610
    1611
    1612
    1613
    1614
    1615
    1616
    1617
    1618
    1619
    1620
    1621
    1622
    1623
    1624
    1625
    1626
    1627
    1628
    1629
    1630
    1631
    1632
    1633
    1634
    1635
    1636
    1637
    1638
    1639
    1640
    1641
    1642
    1643
    1644
    1645
    1646
    1647
    1648
    1649
    1650
    1651
    1652
    1653
    1654
    1655
    1656
    1657
    1658
    1659
    1660
    1661
    1662
    1663
    1664
    1665
    1666
    1667
    1668
    1669
    1670
    1671
    1672
    1673
    1674
    1675
    1676
    1677
    1678
    1679
    1680
    1681
    1682
    1683
    1684
    1685
    1686
    1687
    1688
    1689
    1690
    1691
    1692
    1693
    1694
    1695
    1696
    1697
    1698
    1699
    1700
    1701
    1702
    1703
    1704
    1705
    1706
    1707
    1708
    1709
    1710
    1711
    1712
    1713
    1714
    1715
    1716
    1717
    1718
    1719
    1720
    1721
    1722
    1723
    1724
    1725
    1726
    1727
    1728
    1729
    1730
    1731
    1732
    1733
    1734
    1735
    1736
    1737
    1738
    1739
    1740
    1741
    1742
    1743
    1744
    1745
    1746
    1747
    1748
    1749
    1750
    1751
    1752
    1753
    1754
    1755
    1756
    1757
    1758
    1759
    1760
    1761
    1762
    1763
    1764
    1765
    1766
    1767
    1768
    1769
    1770
    1771
    1772
    1773
    1774
    1775
    1776
    1777
    1778
    1779
    1780
    1781
    1782
    1783
    1784
    1785
    1786
    1787
    1788
    1789
    1790
    1791
    1792
    1793
    1794
    1795
    1796
    1797
    1798
    1799
    1800
    1801
    1802
    1803
    1804
    1805
    1806
    1807
    1808
    1809
    1810
    1811
    1812
    1813
    1814
    1815
    1816
    1817
    1818
    1819
    1820
    1821
    1822
    1823
    1824
    1825
    1826
    1827
    1828
    1829
    1830
    1831
    1832
    1833
    1834
    1835
    1836
    1837
    1838
    1839
    1840
    1841
    1842
    1843
    1844
    1845
    1846
    1847
    1848
    1849
    1850
    1851
    1852
    1853
    1854
    1855
```

Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, P., Couve, E., de Darvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Feihl, H., Portillo, F.G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F., Kurapkut, G., Madueno, E., Maitounam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordstiek, G., Novella, S., de Pablos, B., Perez-Diaz, J.C., Purcell, R., Rammel, B., Rose, W., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J. and Cossart, P. Comparative genomics of *Listeria species* Science 294 (5543), 849-852 (2001)

21537279
11679669

2. (bases 1 to 280050)
Glaser, P., Frangeul, L. and Rusniok, C.
Direct Submission
Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE

E-mail: pglaser@pasteur.fr
Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.

FEATURES

source

1. 280050

/organism="Listeria monocytogenes"

/mol_type="genomic DNA"

/strain="EGD-e"

/db_xref="taxon:1639"

158. .584

/gene="lmo0332"

158. .163

/gene="lmo0332"

172. .507

/gene="lmo0332"

/codon_start=1

/transl_table=11

/protein_id="CAC98411.1"

/db_xref="GI:16409710"

/db_xref="SPTREMBL:O8YA33"

/translation="MIYICALYTFISALVSFGFSLDALLKSRKNGDALINKAYAVS

RSLGLLVAGLFLIFKSDAFVLSLVMIQAQLFDGIGIKISTFKTVGPLLTAAGNV

IMLILFLTI"

565. .584

/gene="lmo0332"

936. .6272

/gene="lmo0333"

936. .6272

/gene="lmo0333"

/note="similar to internalin proteins, putative

peptidoglycan bound protein (LPXTG motif)"

/codon_start=1

/transl_table=11

/protein_id="CAC98412.1"

/db_xref="GI:16409711"

/db_xref="SPTREMBL:O8YA32"

/translation="MKKFSIVIIISVLLGLVLPDFTLLVGADETTVSEDVATKTAEA

DSATEGISETGSDTAEPKEAEAEAREKETTEKEEKAETEPASNIKTEINDKS

QLKQTSKAAVPAGSTYNSLPDNDNLAKLAVITGNAATGNEVDSAALLAISQD

LSGTGNDPTDISNIEGLQYLENTLSLNENNISDLAPLKDLVNLVSLNSRLV

NLGVEDLVNQLVNSANKALEDSQVAPLVLKEISAGQCNKITLLEKNPAGAVLP

ELETFVLEQNDLTNLSLAKLPLKLVYIKGNASLSLETNGATKLQIDASCTDL

ETLGDIGSLELEMQLSGCSKLEITSKLNPLNVIADSCAIEDLGTNLNLENNK

TLVSDNENLTNITATDLPQKTLTLDGGCITSIGTLDNLPKLEKLDLKENQITSIS

EITDPLRLSDVSNVNTLIGDKLPLLEWLVNSNRLSDVSTLTNFPPLNYNIS

NNVIRTKVSTELSKFEFAQNNSISDSIMHDPNLRKVDASNNLTINIGFDNLP

KLQSDVHSNRITSTSVHDLPSLETENAOTNLITGTMDNLPDLTYVNLNFRNIPS

LAPIGDPLNLTIVSDNNSVLSRIGTMGVKPLRIIDLNQNNLYNTGTGECNLSLSD

LTNLTELNNLVVIDISGLSTSLRLIYLNLSNKIEDISALSNLTNLOELTNLENNK

IENISLSDLENLKNLVSKNKIIDISFVANNVRGAIVTASQTYTLPTVLSYQSSP

TIDNPVIWYDGTLLAPSSIGNSGNYKDKITWTNMTATSSSTLPFNRLKDGTLFSGT

VTPYKAAKVTADAQTYTIGDTISEEQFLKQVNAKSDGAPVSTFATVDLNTFG

EYEVTLTSEKDIGQSKVIVKLVHGAPVISADQTIISYDKHATITTEKQLEDIHAST

DLDTAITTNFSTAVLNKGGDYTVALNSENEDGVKAETVTVTVTVNKKDPAIISAKTE

ITYDKFSKTEAFLDDIDADTNDGSI VTSNFAFATVNLKAGDYTVTLNSINSDGVA
TPTAIIVHVEKIKIATISNTAQQYKAKINETQFLKDVHASINASPITVAVLESDFE
TVKLDVPCQTVTVTATNEDGVSAPKEVSVIRKIPAPEITADKEITYPKFDEVSE
AEFLNDIHATISDKVAITSNSTVNLNKGDTVTLNATNEDGVKATPVEIVHVQ
QSERPVITADATISVDKFNITEAKFLEDIHATSDGQSSSTVITSNFQATNFKTAMS
YTVTLNADNEDGISAEPAVTVTINKPEAALKADEAVSYAKNEAVTESDFFKDVHLE
GTEATSKATKSTFDSVVRDSTGDTVTINATNEDGAVSTPIEVIVHIEASAPVIT
ANAEVKNKHEQTDERRFLYDSEAKIDEANVEIKTDFAEKVDINKVGTVTLTATNE
DQCANPVEVSVISDAAAEKNVYVDENGSEISAAETLTGNLDETFSIDAKSIAGY
KCDATLSGVFSTVEQTVVPHYKAIKPGVVTIKYEDTNGKAVDEKQITGEVGDPRAE
AQTVSGISCRATASGRITTEPQTITFTYTSATPSSKSGEITVQYVYDESKKLADSKV
TGNISDTSYSEAKAIEGYSVGDGSAKVGFTEKSTQVTFPKMKTQVSKDDPKRGKT
NQPSSTDTKLKNNLSLPATGDTENMLAVLIGFNNMLIVASIFLFRPKPTNQ"

gene

/gene="lmo0334"

6491. .7032

/gene="lmo0334"

6491. .6496

/gene="lmo0334"

6508. .7032

/gene="lmo0334"

/codon_start=1

/transl_table=11

/protein_id="CAC98413.1"

/db_xref="GI:16409712"

/db_xref="SPTREMBL:O8YA31"

/translation="MNKIIYSQYTSQYATELTSQKKAIPMYSVESIPESIVAEDE

IIVGSIYMGKVLGPFDKFSKYKTLSSKLIVISAGYDPTKENTDNIELTVKENLQK

TNFLKDIPLCLAGKLDTQQLRKLKILNALYKSARKKESDLNDRDVRAYEDDS

QIDLNLKLDGVDRL"

join(7235. .7240,7248. .7532)

/gene="lmo0335"

7235. .7240

/gene="lmo0335"

7248. .7532

/gene="lmo0335"

/codon_start=1

/transl_table=11

/protein_id="CAC98414.1"

/db_xref="GI:16409713"

/db_xref="SPTREMBL:O8YA30"

/translation="KREIKVDERTFOQHATKLASESTGSLVPLKNGNMAYSKANSDIQ

LRSALIELVDVVVEDFOHVTKNSRUKRGGIAYAKOQDLMGKINOLEVR"

7519. .7524

/gene="lmo0336"

7519. .7524

/gene="lmo0336"

7533. .7901

/gene="lmo0336"

7533. .7901

/gene="lmo0336"

/codon_start=1

/transl_table=11

/protein_id="CAC98415.1"

/db_xref="GI:16409714"

/db_xref="SPTREMBL:O8YA29"

/translation="MDSAHSOLEQQLQKIKAKITAEADQTRRKQNEQDWLEEDSN

QLTQEKRVLLDFLRSGMQCEBASSPHRYLEEQQHEBSQVWRDLQDKRSLDTELOEN

KORLHLEAKQATLQKENSQ"

7885. .7890

/gene="lmo0337"

7885. .7890

/gene="lmo0337"

7898. .9373

/gene="lmo0337"

7898. .9373

/gene="lmo0337"

/codon_start=1

/transl_table=11

/protein_id="CAC98416.1"

/db_xref="GI:16409715"

/db_xref="SPTREMBL:O8YA28"

/translation="MSRIDIAELNDFLHGLRSNAEAKAMIRKIKEAMMDYADNRK

GEAVTSVKQFSGSTYISICQSIIEALDESEERLAQIRFSGVSDSPSPARIDAEILQ

EAMKVSQQRKEEDLHRQLTAPNTKPDQOVYAVKRSVHTQLLKAIBOENILEKYL

```
AFQSHGQFPSSALDELIRATARAQVQLHHVSNFNDKTYTSVVPKSAANSULLMKKARD
KARKENDQPKPAFEDYTVLAYTYNDQGETVWMLLEKDKRKRVENKLODFLEKHG
QELDPLLYTNLSSEELERKVNDSWKEGYNLNGQKVSQVSGATLRSSAYVASKMOWD
DAGLTDMALGFGIAAARNKPKVKEKLNPAETKQTVNGEKWNNYFKKYKAGNVOWK
PSSFEDILINPERLYGSTKKEVPQMLGEGWKEETYSAGNGWKFTNEGDGMPYHPGE
GIHKGSYGFPSSDGTQKVLVKGNDYIDFSKDKATIIIFKGE"
gene      /gene="lmo0338"
RBS       9362..9368
          /gene="lmo0338"
gene      9377..9757
          /gene="lmo0338"
CDS       9377..9757
          /codon_start=1
          /transl_table=11
          /protein_id="CAC98417.1"
          /db_xref="GI:16409716"
          /db_xref="SPTREMBL:O8YA27"
          /translations="MKKLTVGGKFDYIDTLEKSGMFILELSNEEIEETIFEDPIVGV
TSFSSKNLSSELKANIEIDDEMEKNAYLLREKVLKIDNTDLWNINSVRQSSSEWLDIFR
RSDCLKRQLKERSDEIEVYKTL"
gene      join(10019..10024,10032..10403)
          /gene="lmo0339"
RBS       10019..10024
          /gene="lmo0339"
CDS       10032..10403
          /gene="lmo0339"
          /note="weakly similar to inorganic pyrophosphatase"
          /codon_start=1
          /transl_table=11
          /protein_id="CAC98418.1"
          /db_xref="GI:16409717"
          /db_xref="SPTREMBL:O8YA26"
          /translations="WVSRKQSIQKQLSKTKQTLNWTIDRPIGYTDKFGNVYPLNYG
FVEGVIGGGSEQDVVLSRKISEPIDTFQGDLIIVITRNDVDEKWWVAAPNENFTI
EETREKVHIEQVDFSEIKLI"
gene      join(10402..10407,10416..10634)
          /gene="lmo0340"
RBS       10402..10407
          /gene="lmo0340"
CDS       10416..10634
          /gene="lmo0340"
          /codon_start=1

Query Match      6.5%; Score 127.8; DB 1; Length 280050;
Best Local Similarity 51.3%; Pred. No. 4.4e-18;
Matches 325; Conservative 0; Mismatches 302; Indels 6; Gaps 1;

QY 96 TTGATTACGGTGTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGCACACTTTGGATTG 155
DB 243522 TTGACTATGATACAGGAAACACGAGNAGCATCAGCAAGCACTTGATTTATCGACTAC 243463

QY 156 ATATCAAGAATGTGCAAAAGCCAGAGATATCTAAATGCTAAGCGCCTTATCTTTCCCTG 215
DB 243462 AAAACAAAATTTCTAGTGATGCAACAGAAATCTCACAGGCTGACGGGTTATTTTACCAG 243403

QY 216 GCCTTGGGCGCTTGCACCTGCATGATGTTCTTATTCTGTAAGGCTGGCTGAAGCAC 275
DB 243402 GAGTTGGCGCTTATCTCTGAAGCTATGAAAGAACTCACTCGCGCGGGCTAAATAAACAT 243343

QY 276 TCTGTACTTACATTCAGAATGATCGACCTTCTCTGGGTATATGCTGGGATTCGACGTAC 335
DB 243342 TAAAGNAATCGTCCGGTGGCAACCAATTTCTGTGTGTTCTGTGTATGCACTTT 243283

QY 336 TCTTTGAGTCAAGTGAAGAAATAGTCCAAATCAAGTCTTGGCTTGATCTCGGACGGG 395
DB 243282 TACTCGAATCAAGTAACGACATAGCTTTTACTAATGCTCTCGGTCTCATTTCCCGGTCA 243223

QY 396 TTGGCGCTTTTGATCATCCAAATGGTTTAAGGGTGCCACATATTTGATGCGATGCGCTGG 455
DB 243222 TCGAAAAATTCACGAAGATCCCGGAATTTGCCGTTCCACATATGGGCTGGAAATCAACTAC 243163
```

```
QY 456 ATATAAAGGAAGGTCAGCAATTTTAGATGATGTGGGGAATCAACATGTGATTTTGTTC 515
DB 243162 AAATCAACGTCGCCACCCCGCTTACAAAAACAACCTTGATGGAGAATACGTTTATTACGTT 243103

QY 516 ACTCATATCGAGCCATGCGGAGGACAAACAAGAGTGGATTTTCATCTACATCGAGTATG 575
DB 243102 ACTCTTATTACGCCAATTCGCCAG-----AAGAATACATATATCGCAACAAGTGGCTATT 243049

QY 576 GTGACGATTTTATTCATCCATTCAGAGGGAAATGTTTCATGCACTCCCAATTTTCATCCCG 635
DB 243048 CTATTGAAGTTCTTGGCATATAATAAGGCNATATTTACGGCGCCAGTTCCATCTCG 242989

QY 636 AGAAGAGTCGAGGTGTGGACTTTTCATATTGAGAAGATTTTGAATGCTGATCTCTTTA 695
DB 242988 AAAAAAGCGCCAAATCGGCTTTGAAATTTTAAAGGATTTAAGGAGGTCACTTATTCAT 242929

QY 696 ACAACAAGACAGAGCCCAATGAATGGAAGG 728
DB 242928 GCAAAATCTTCCCGAGCAATAGATTTTAAAAAATGG 242896

RESULT 15
AX641665/c 349980 bp DNA linear PAT 21-FEB-2003
LOCUS      Sequence 2855 from Patent WO0101118.
DEFINITION AX641665
ACCESSION  AX641665
VERSION     AX641665.1 GI:28474426
KEYWORDS    Listeria monocytogenes
SOURCE       Listeria monocytogenes
ORGANISM     Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE    1
AUTHORS      Buchrieser, C., Frangeul, L., Couve, E., Rueniok, C., Fsihi, H.,
              Dehoux, P., Dussange, O., Chetouani, F., Nedjari, H., Glaeser, P.,
              Kunst, F., Cossart, P., Daniels, J., Goebel, W., Kreft, J., Kuhn, M.,
              Ng, E., Vazquez-Boland, J., Dominguez-Bernal, G., Garrido-Garcia, P.,
              Tierrez-Martinez, A., Amend, A., Chakraborty, T., Domann, E., Hain, T.,
              Berche, P., Charbit, A., Durant, L., Perez-Diaz, J. C., Baquero, F.,
              Garcia del Portillo, F., Gomez-Lopez, N., Madueno, E., de Pablo, B.,
              Wehlund, J., Kaerst, U., Entian, K. D., Hauf, J., Rose, M. and Voss, H.
              Listeria monocytogenes genome, polypeptides and uses
              Patent: WO 010118-A 2855 11-APR-2001;
              INSTITUT PASTEUR (FR)
FEATURES     Location/Qualifiers
              1..349980
              /organism="Listeria monocytogenes"
              /mol_type="genomic DNA"
              /db_xref="taxon:1639"
              /note="Seq 1 for big: original length: 2944528-splitted
              into 10 <223>seq 0001 : From 0.000.001 to 0.349.980
              length: 349.980 <223>seq 2855 : From 0.300.001 to
              0.649.980 length: 349.980 <223>seq 2856 : From 0.600.001
              to 0.949.980 length: 349.980 <223>seq 2857 : From
              0.900.001 to 1.249.980 length: 349.980 <223>seq 2858 :
              From 1.200.001 to 1.549.980 length: 349.980 <223>seq 2859
              : From 1.500.001 to 1.849.980 length: 349.980 <223>seq
              2860 : From 1.800.001 to 2.149.980 length: 349.980
              <223>seq 2861 : From 2.100.001 to 2.449.980 length:
              349.980 <223>seq 2862 : From 2.400.001 to 2.749.980
              length: 349.980 <223>seq 2863 : From 2.700.001 to
              2.944.528 length: 244.528"
BASE COUNT  115068 a 60319 c 71179 g 103414 t
ORIGIN
```

Query Match 6.5%; Score 127.8; DB 6; Length 349980;
Best Local Similarity 51.3%; Pred. No. 4.4e-18;
Matches 325; Conservative 0; Mismatches 302; Indels 6; Gaps 1;

```
QY 96 TTGATTACGGTGTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGCACACTTGGATTG 155
DB 303522 TTGACTATGATACAGGAAACACGAGAGCATCAGCAAGCACTTGATTTTATCGGACTAC 303463

QY 156 ATATCAAGAATGTGCAAAAGCCAGAGGATATCTAAATGCTAAGCGCCTTATCTTTCCCTG 215
```

```
Db 303462 AAAACAAATTTCTAGTGATGCAACAGAAATCTCACAGGCTGACGGGGTTATTTTACCAG 303403
Qy 216 GCGTTGGGGCTTTTGCACTGCAATGATGTTCTTATTCGTAAAGGCTGGCTGAAGCAC 275
Db 303402 GAGTTGGCGCTTATCTCTGAAGCTATGAAGAACTCACTCGCGCGGGCTAAATAAACAT 303343
Qy 276 TCTGTACTTACATTCAGATCGACCTTTCCTGGGTATATCGCTGGGATTCAGCTAC 335
Db 303342 TAAAGAAATCGCTGCGCGGTGGCAACCAATCTTGGTGTTCCTGGGTATGCACTTT 303283
Qy 336 TCTTTGAGTCAAGTGAAGAAATCGTCCAATTCAGGTCTTGGCTTGATTCCTGGACGG 395
Db 303282 TACTCGAATCAAGTAAAGACATAGCTTACTAATGTCTCGGTCTCATTTCCCGGTATG 303223
Qy 396 TTGGCGCTTTTGAATCATCCAAATCGTTTAAAGGGTCCACATATTTGGATGGGATGCCCTGG 455
Db 303222 TCGAAAAATTACCAGAAAGATCCCGAATTTGCCGTTCCACATATGGGCTGGAATCAACTAC 303163
Qy 456 ATATAAGGAGGCTCAGCAATTTTAGATGATGTGGGAATCAACATGCTGATTTTGTTC 515
Db 303162 AAATCNAACGTCACCCCGCTTACAAACAACTTGATGGAGAAATCGTTTATACGTTT 303103
Qy 516 ACTCATATCGAGCCAATGCCGAGGACAAACAAAGAGTGGATTCATCTACATGCGACTATG 575
Db 303102 ACTCTATTAGCCCAATTTGCCAG-----AAGATACATTTATCGCAACAGTGGCTATT 303049
Qy 576 GTGACGATTTTATGTCATCCATTCAGAGGGAATGTTTCATGCAATCCCAATTTTCATCCCG 635
Db 303048 CTATTGAAGTTCTCGGCATGATAAATAAAGGCAATATTACGGCGCCAGTTCATCCTG 302989
Qy 636 AGAAGAGTGGAGGTGTTGGACTTTCATATGAGAAGATTTTGAATGCTGATTCCTTTA 695
Db 302988 AAAAAAGCGCCAAATCGGTCTTGAAATTTTAAAGGATTTAAGGAGGTCACTTATTCAT 302929
Qy 696 ACAACAAAAGACAGAAGCCAAATGAATGGAAGG 728
Db 302928 GCAATCTTCCAGCAATAGATTTAAAAATGG 302896
```

Search completed: January 20, 2004, 23:49:58
Job time : 7204 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 18:26:05 ; Search time 535 Seconds
(without alignments)

9869.347 Million cell updates/sec

Title: US-09-831-233A-1

Perfect score: 1956

Sequence: 1 gcacgagaaggaactggca.....aaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1956	100.0	1956	21 AAA27325	Impatiens balsamia
2	521.6	26.7	1084	21 AAA27327	Corn putative glut
3	266.6	13.6	671	21 AAA27331	Soybean putative g
C 4	127.8	6.5	2944528	24 ABA03041	Listeria monocytog
C 5	125.4	6.4	1163020	24 ABQ67197	Listeria innocua c
C 6	125.4	6.4	3011208	24 ABQ69245	Listeria innocua D
7	117.2	6.0	4584	24 ABQ70963	Listeria monocytog
8	113.2	5.8	1860	21 AAF08042	Fusarium venenatum

9	104	5.3	178	21 AAA27330	Soybean putative g
10	99.2	5.1	476	25 AB255560	Aspergillus oryzae
C 11	98.6	5.0	1664976	19 AAV21209	Methanococcus jann
12	91.2	4.7	1523	24 ABO69172	Listeria monocytog
13	91.2	4.7	2365589	24 ABA90521	Genomic sequence o
14	87.6	4.5	564	24 ABK77558	Bacillus clausii g
15	85.6	4.4	777	24 ABN92347	Staphylococcus epi
16	84.4	4.3	405	24 ABK77570	Bacillus clausii g
17	83.6	4.3	757	24 ABK73011	Bacillus lichenifo
18	83.6	4.3	757	24 ABK73180	Bacillus lichenifo
19	82.6	4.2	636	24 ABK73073	Bacillus lichenifo
20	80.8	4.1	744	22 AAH52449	S. epidermidis ope
21	80.8	4.1	3022	22 AAH54365	S. epidermidis gen
22	77.2	3.9	344	24 ABN77661	Human ORF2608 cDNA
23	75	3.8	6346	18 AAV74548	Staphylococcus aur
24	74.8	3.8	611	21 AAF08136	Fusarium venenatum
25	74.2	3.8	587	21 AAA27326	Corn putative glut
26	71.4	3.7	49646	21 AAA81457	N. meningitidis pa
C 27	71.4	3.7	349980	21 AAF21608	Neisseria meningit
C 28	71.4	3.7	1437668	21 AAA81490	N. meningitidis B
C 29	69.8	3.6	494	23 ABV10021	Human prostate exp
30	69.8	3.6	495	21 ABA27329	Rice glutamine ami
31	68.4	3.5	774	22 AAH67256	C glutamicum codin
32	68.4	3.5	897	22 AAF71947	Corynebacterium gl
C 33	68.4	3.5	349980	22 AAH68531	C glutamicum codin
C 34	68.4	3.5	1830121	17 AAT42063	Haemophilus influe
C 35	68.2	3.5	738	25 ABZ38917	N. gonorrhoeae nuc
36	67.8	3.5	765	25 ABZ38918	N. gonorrhoeae nuc
37	66.8	3.4	24417	18 AAT97221	Pseudomonas aerugi
C 38	65.8	3.4	556	23 ABV40063	Human prostate exp
C 39	65.8	3.4	556	23 ABV40163	Human prostate exp
C 40	65.8	3.4	556	23 ABV42105	Human prostate exp
C 41	65.8	3.4	556	23 ABV43601	Human prostate exp
C 42	65.4	3.3	64081	24 ABK92787	Buchnera sp. genom
C 43	65	3.3	309	23 ABK44994	Human prostate exp
C 44	64.2	3.3	419	25 ABK46069	Bovine EST associa
C 45	63.6	3.3	100848	22 AAF28552	Genomic fragment #

ALIGNMENTS

RESULT 1
AAA27325
ID AAA27325 standard; cDNA; 1956 BP.
XX
AC AAA27325;
XX
DT 08-AUG-2000 (first entry)
XX
DE Impatiens balsamia glutamine amidotransferase gene.
XX
KW Glutamine amidotransferase; histidine biosynthesis; herbicide;
KW fungicide; ss.
XX
OS Impatiens balsamia.
XX
FH Key Location/Qualifiers
FT CDS 2..1678
FT /tag= a
FT /product= "glutamine amidotransferase"
FT /partial
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

QY 1801 ATTATTTGGCAATTTGTTTGAATAGATTTGAGTTTTTAGACCTTGGTGTGCTGTTT 1860
DB 1801 ATTATTTGGCAATTTGTTTGAATAGATTTGAGTTTTTAGACCTTGGTGTGCTGTTT 1860
QY 1861 TATCTCTAGGCCATGTTTGTGGATTATATACAGTGTGAATTAATAATAATGCGTA 1920
DB 1861 TATCTCTAGGCCATGTTTGTGGATTATATACAGTGTGAATTAATAATAATGCGTA 1920
QY 1921 TGAATTTATGCTTTTAAAAAATAAAAAAATAAAAAA 1956
DB 1921 TGAATTTATGCTTTTAAAAAATAAAAAAATAAAAAA 1956

RESULT 2

AAA27327

ID AAA27327 standard; cDNA; 1084 BP.

XX

AC AAA27327;

XX 08-AUG-2000 (first entry)

XX Corn putative glutamine amidotransferase gene 2.

XX Glutamine amidotransferase; histidine biosynthesis; herbicide;
KW fungicide; corn; ss.

XX

XX Zea mays.

XX

XX Key Location/Qualifiers

FT CDS 2..889

FT /*tag= a

FT /product= "glutamine amidotransferase"

FT /partial

XX

XX W0200028053-A2.

XX

XX 18-MAY-2000.

XX

XX 04-NOV-1999; 99WO-US25950.

XX

XX 05-NOV-1998; 98US-0107275.

XX

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX

XX Allen SM, Huang LL, Falco SC, Rafaleki AJ;

XX

XX WPI; 2000-376564/32.

XX

XX P-PSDB; AAY94226.

XX

XX Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
PT primers -

XX

XX Claim 2; Page 43; 53pp; English.

XX

XX The present sequence is a Zea mays (corn) putative glutamine
CC amidotransferase coding sequence. The protein forms a crucial stage
CC in the histidine biosynthesis pathway. The sequence can be used to
CC create transgenic plants which express different amounts of the protein,
CC to identify loss of function mutants and to produce the protein in a host
CC cell, for example a bacterium. The protein can be used to identify
CC inhibitors which may be useful as fungicides and herbicides. The gene
CC was identified by computer screening for sequences which could encode
CC histidine biosynthetic enzymes.

XX

SQ Sequence 1084 BP; 316 A; 182 C; 281 G; 305 T; 0 other;

XX

Query Match 26.7%; Score 521.6; DB 21; Length 1084;

Best Local Similarity 73.6%; Pred. No. 2.1e-107;

Matches 665; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

XX

QY 791 CTTGTTGTTACCAAGGAGACCAATATGATGTGAGAGAACGTACAGAGAGATGAGTC 850

DB 2 CTAGTGGTAACAAAGGCGATCAATATGATGTAAGAGATCATACTAGCAGCAAGAGGTA 61
QY 851 AGAAACCTTGGCAAGCCTGTGTAACCTTGTGGGAGTATTTATTAGACGGTCTGATGAG 910
DB 62 AGAAACCTTGGCAAGCAGTTCGATTTAGCAAGCAGTACTACATAGACGGTCTGATGAG 121
QY 911 GTCAGCTTCTTAAACATTACTTGGTTTTCCGGGACTTCCCTCTAGGCGCATCTACCATGCTA 970
DB 122 GTCAGCTTCTTGAATATAAATGCTTTCCGCTGACTTTCCATTGGGTGATTTGCCAATGCTA 181
QY 971 CAGGCTTTCGACCGCATCTGAAACGTTTTTGTGCCATTAATCTGTGGGGGTGGCATC 1030
DB 182 GAGGTACTGCGTTGTGCTCTCTGAAAAGGTTTTTGTGCCACTTACAGTTGGTGGGCGATA 241
QY 1031 AGGGATTTTACTGATGCAATGGAAGTATTCTTAGTCTAGAGTGGCTTCAGAGTAT 1090
DB 242 CGAGACTTCACAGTGAATGGAAGTACTTCTCAAGTTTGGAGGTAGCATCAGAAATAT 301
QY 1091 TTCAGATCGGGCGCCGATTAAGGTTTCGATCGGAAGTATGATGATGATGATGATGATGAT 1150
DB 302 TTCAGGTCGGTGTCTGACAAAATTTCAATTGGAAGTATGCTGTTTATGCTGCTGAAGCC 361
QY 1151 TATATTAAACCCGGAGTGAAGACAGAAAGAGCAGCATAGAGCAGATATCTACAGTATAT 1210
DB 362 TTTTACAGACTGGTGTAAAGACAGGGAAGCAGCTTGGAGCAAAATCTCTAGAGTATAT 421
QY 1211 GGTAAACGAGCGAGTGGTGTAGCATTGATCTCCGCGGAGTTTACTTGAAGAAACCCGAT 1270
DB 422 GGCATTCAGGCTGTAGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 481
QY 1271 GAAGTAGAATTTAAAGCCATCAAGTAAGCCATCCAGAGTCCAAACGGTGAAGGATATGCC 1330
DB 482 GATGTCCATTTAAACTGTAAAGTGTCCACTAAAGTCCATCCGGAGAGAGATATGCA 541
QY 1331 TGGTATCAGTGCATCTTTAATGTTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1390
DB 542 TGGTACAGTGCACAGTGAATGTTGAGCGTGTAGCGAGCTATAGGAGCATATGAACATA 601
QY 1391 GCTAAGGCTTTGAGGAACTTGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1450
DB 602 GCGAAAGCTGTGGAAGAAATGGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
QY 1451 GGTCAAGGAAAGAGATTTCGATATAGATCTGATCAAGCTAATATCCGATGCTGTGAACAT 1510
DB 662 GCGCAAGGTTGTGGATTGACATAGATTTGGTTAAATGGTTCTGATGCTGTGACATC 721
QY 1511 CCGTTTATCGCAAGCAGCGGTGCGAGAGTGGTGTGATCATCTTCCGAAAGTCTTTTAATGAA 1570
DB 722 CCGTTCATTGCGAGCAGTGGTGTGAGCTGTTCAACATTTTCTGAAATTTTGTGAGAAA 781
QY 1571 ACCAAGCATCTGCTGCCCTTGCAGCTGGCATTTTCCATCGCAAGAGGTTCCCAATTAAG 1630
DB 782 ACAATGCTTCTGCTGCCCTTGCAGCTGGCATTTTCCACCGGAAAGAGGTTCTCTACTA 841
QY 1631 GCTGTTAAAGAGCATTGTTGAAGGAGGATTGAAGTTAGATTGTAAGCGGAGATCAC 1690
DB 842 GCAGTGAAGAGCATCTGCTCAATGCTGTGTGGAGTCAAGGTTGTAACAGGAGATCCT 901
QY 1691 TTGG 1694
DB 902 TCG 905

RESULT 3

AAA27331

ID AAA27331 standard; cDNA; 671 BP.

XX

AC AAA27331;

XX

DT 08-AUG-2000 (first entry)

XX

DE Soybean putative glutamine amidotransferase gene 2.


```
Db 511852 CAAACGAATACATTATTGCAACTAGCGGTTATTCTGTTGAAGTTCCAAAGTATGATAAATA 511793
Qy 603 AGGAAATGTTTCATGAGTCAATTTTCATCCGAGAAAGATGGAGGTGTTGGACTTTCCA 662
Db 511792 AAGCAATATTACGGCGCCAGTTTCCACGAGAAAAAGCGGTCAAAATCGGGCTTGAAA 511733
Qy 663 TATTGCAAGATTTTTCGAATG 683
Db 511732 TTTTAAAGGGTTTAAAGGAG 511712

RESULT 6
ABQ69245/c
ID ABQ69245 standard; DNA; 3011208 BP.
XX AC
XX AC
XX AC
XX 29-AUG-2002 (first entry)
XX DE
XX Listeria innocua DNA sequence #684.
XX KW
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX KW
XX infection; ds.
XX OS
XX Listeria innocua.
XX XX
XX WO200228891-A2.
XX PN
XX 11-APR-2002.
XX PP
XX 04-OCT-2001; 2001WO-FR03061.
XX PR
XX 04-OCT-2000; 2000FR-0012697.
XX PA
XX (INSP ) INST PASTEUR.
XX PA
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX FI
XX Kunst F, Glaser P;
XX XX
XX WPI; 2002-332479/37.
XX XX
XX New genomic sequences from Listeria species, useful for detection,
XX PT
XX treatment and prevention of infection, also related polypeptides,
XX PT
XX antibodies and modulators
XX XX
XX Claim 5; SEQ ID 2058; 180pp; French.
XX XX
XX The present invention relates to nucleic acid sequences
XX CC
XX (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
XX CC
XX and primers for identification and/or detection of Listeria (e.g. as
XX CC
XX contaminants in foods, or mutational analysis) and for analysis of
XX CC
XX gene expression. Proteins encoded by the nucleic acid sequences can be
XX CC
XX used to screen for compounds that modulate gene expression, replication
XX CC
XX and pathogenicity of Listeria (potential therapeutic agents), also for
XX CC
XX treating infections by Listeria, and are useful as immunogens in
XX CC
XX anti-Listeria vaccines.
XX CC
XX Note: The sequence data for this patent did not form part
XX CC
XX of the printed specification, but was obtained in electronic format
XX CC
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;
SQ
Query Match 6.4%; Score 125.4; DB 24; Length 3011208;
Best Local Similarity 51.4%; Pred. No. 2.3e-17;
Matches 319; Conservative 0; Mismatches 296; Indels 6; Gaps 1;

Qy 63 CTGTGTCAGCGGAGATTCTGTGTGACTTTTACTTGTATTGATTCGGTCTGGAAATGTTAGGA 122
Db 602591 CTACTAAGAGAGTCCTGTAAATGATGTTATATTGATATGATACAGAAACACGAGA 602532
Qy 123 GTGTGAGGAGCGCATCCGACACTTGGATTGTATCAAGATGTGCAAAAGCCAGG 182
Db 602531 GCATCAGCAAGCATTGATTTTATTTGCTCTTCAAAATAAAATTTCTAGCAACCCGCGG 602472
```

```
Qy 183 ATATTTCTAAATGCTAAGCGCCTTATCTTTCTCGGCGTTGGGCGCTTTGACCTGCAATGG 242
Db 602471 AAATTTTGAAGCGGATGGCGTTATTTTACCTGGAGTCGGAGCTTTCCCGAAGCGATGC 602412
Qy 243 ATGTCTTTATTCTGAAGGGCTGGCTGAAGCACTCTGTACTTACATTCAGATGATCGAC 302
Db 602411 AGGAACCTAGGTCCGCGCGTCTGGATGTGACGTAAAGAAATGGCAGAATCTGGCAAAAC 602352
Qy 303 CTTTCTCGGTATATGCCCTGGGATTCAGCTACTCTTTGAGTCAAGTGAAGAAATGGTTC 362
Db 602351 CGATGCTCGGTGTTGTTCTTGGTATGCAACTGCTACTTGAATCAAGTGACGAACATCTCT 602292
Qy 363 CAATTCGAAGTCTTGGCTTTGATTTCTCGGACGGGTTGGCGGTTTGTGAATCATCCAAATGGTT 422
Db 602291 TTACTAAAGACTCGGTCTCATTTCCAGGACATGTGCAAAAACCTACCAGACGAACCTGGAT 602232
Qy 423 TAAGGTGCCACATATTGGATGGCATGCTTGGATATAAAGGAAGGTCAGCAATTTTAG 482
Db 602231 TTGCGGTTCCGCACATGGCTGGAATCAACTTGAATAAAACGACGACCCCACTCACTA 602172
Qy 483 ATGATGTGGGAATCAACATGTGTATTTTGTTCACCTCATATCGAGCCAATGCCGAGACA 542
Db 602171 AAAAATCGCGCGTGAATATGCTATTACGTCCATTCTTA-----CTATGCAAAATGTC 602118
Qy 543 ACAAGAGTGGATTTTCATCTACATGAGCTATGTCAGCATGTCGTCGATTTTATTTGTCATCCATTGAGA 602
Db 602117 CAACGAAATACATATTGTCACATGAGCGGTTATTCCTGTAAGTTCCAGATGATGATAAATA 602058
Qy 603 AGGAAATGTTTCATGAGTCCCAATTTTCATCCGAGAAAGATGGAGGTGTTGGACTTTTCCA 662
Db 602057 AAGGCAATATTTACGGCGCCAGTTTTCACCCAGAAAAAGCGGTCAAAATCGGCGCTTGAAA 601998
Qy 663 TATTGAGAAGATTTTGAATG 683
Db 601997 TTTTAAAGGGTTTAAAGGAG 601977
```

```
RESULT 7
ABQ70963
ID ABQ70963 standard; DNA; 4584 BP.
XX AC
XX AC
XX AC
XX 29-AUG-2002 (first entry)
XX DT
XX DE
XX DE
XX Listeria monocytogenes 4b contig DNA sequence #905.
XX KW
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX KW
XX infection; ds.
XX OS
XX Listeria monocytogenes 4b.
XX XX
XX WO200228891-A2.
XX PN
XX 11-APR-2002.
XX PD
XX XX
XX 04-OCT-2001; 2001WO-FR03061.
XX PF
XX 04-OCT-2000; 2000FR-0012697.
XX PR
XX (INSP ) INST PASTEUR.
XX PA
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX XX
XX Kunst F, Glaser P;
XX XX
XX WPI; 2002-332479/37.
XX DR
XX New genomic sequences from Listeria species, useful for detection,
XX PT
XX treatment and prevention of infection, also related polypeptides,
XX PT
XX antibodies and modulators
XX XX
XX Claim 14; SEQ ID 3776; 180pp; French.
XX PS
```

XX The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ671212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4584 BP; 1452 A; 1000 C; 914 G; 1212 T; 6 other;

Query Match 6.0%; Score 117.2; DB 24; Length 4584;
Best Local Similarity 51.0%; Pred. No. 2.2e-16;
Matches 306; Conservative 0; Mismatches 288; Indels 6; Gaps 1;
QY 84 TGGTGACTTTACTTTGATTACGGTCTGCGAATGTTAGGAGTGTGAGGAAGCCATCGCA 143
DB 1544 TGAATGTTATTTGACTACGATACAGGAATACGAAAGCATCAGTAAGCCCTTGATT 1603
QY 144 CACTTGGATTGTATATCAAGATGTGCAAAAGCCAGAGGATATTTCTAAATGCTAAAGGCC 203
DB 1604 TTATCGGGTTACAAAACAAAATTTCTAGTGATAAAACAGAAATCGCTCAAGCTGATGGTG 1663
QY 204 TTATCTTTCTGGGTTGGGCTTTGCACCTGCAATGGATGTTCTTATTTCGTAAGGGC 263
DB 1664 TTATTTTGGCCGAGTTGGTCTTATCCAGAGCCATGCAAGAACTCACTCGACGGGAT 1723
QY 264 TGCTGAAGCACTCTGTTACTTACATTCAGAAATGATGCACTTCTTGGGTATATGCTGG 323
DB 1724 TAGATAAAACATTAAGAAATTTGCTACCCCTGGTAACCAATTTCTGGGTTTGCTTG 1783
QY 324 GATTGCACTACTCTTTGAGTCAAGTGAAGAAATGCTCCAAATTCAGAGTCTTTGGCTGA 383
DB 1784 GTATGCAACTATTACTCGAATCAAGTAACGAACATAGCTATCTAAAGGGCTCGGCTCA 1843
QY 384 TTCTGTGACGGGTTGGCGTTTGAATCATCTCCAAATGTTTAAAGGTGCGCACATTTGGAT 443
DB 1844 TTCCCGGTCTGATCGAAATGCTACCAAGCAATCCGAATTTGCGGTTCCGCACATGGCT 1903
QY 444 GGCATGCTTGGATATAAGAGGGTCAAGCAATTTTATAGATGATGTTGGGAAATCAACATG 503
DB 1904 GGAATCAATTAACAATCAAGCAACCGCCGCTTACACAAACATCGCTGGAGAAATATG 1963
QY 504 TGTATTTTGTCTACTCATATCGAGCCAAATCCGAGGACACAAAGAGTGGATTTCATCTA 563
DB 1964 TCTATTACGTTTCATCTCTA-----CTATGCCAACTGTCCAGAGCTTACATTTATCGCA 2017
QY 564 CATGCACTATGTTGAGCATTTTATTCATCTCCATTCAGAAAGGAATGTTTCATGCAATCC 623
DB 2018 CAAGTGGATATGCTATCGATATCCCTGGCATGATAAATAACGGACATATTTACGGAGCAC 2077
QY 624 AATTTCAATCCGAGAGAGTGGAGTGTGGAGTCTTCCATATTCAGAGATTTTGAATG 683
DB 2078 AGTTGCATCTCTGAAAGAGCGGCAATCGTCTTGAATTTTAAAGGTTTAAAGAGG 2137

RESULT 8

AAF08042
ID AAF08042 standard; cDNA; 1860 BP.
XX
AC AAF08042;
XX
DT 13-MAR-2001 (first entry)
XX
XX Fusarium venenatum EST SEQ ID NO:565.
DE Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW

Query Match 5.8%; Score 113.2; DB 21; Length 1860;
Best Local Similarity 52.1%; Pred. No. 1.3e-15;
Matches 244; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
QY 1307 GGTCCAAACGGTGAGGAATATGCTGGTATCATGTCACATGTTTAAATGGTGACAGAGGG 1366
DB 2 GGTCTTAAGGGGAGGAGTACTGTTGGTNCNCCTGCACAAATCAAGGGTGGCCGGAAGT 61
QY 1367 AGACCCATCGAGCTTATGAATAGCTAGGGTGTGAGGAACTTCGAGCTCGAGAAATA 1426
DB 62 CGGGATATGGATGTAGTGGAGCTTGTGCAANCGTTCGAGCGGATGGGAACCGGANAAGATT 121
QY 1427 TTATTGAATGCAATTCATTTGATGTGTCAAGGAAAGAGGATTCGATATAGATCTGATCAAG 1486
DB 122 CTCCTCACTGCAATTCAGCAAGGACGACCAACAGTGGGTTGACTTCGAGCTCATCAAC 191
QY 1487 CTAATATCCGATGCTGTGAACATTTCTGTTATCGCAAGCAGCGGTGACAGAGTCTGAT 1546
DB 182 CAGGTCAAGGGAGCTGTCAAGATCCCCTCATCGCTCAAGTGGTGGTCTGCGCAACCCGGCT 241

XX DE Bacillus clausii genomic sequence tag (GST) #401.
 XX KW Differential gene expression; genomic sequenced tag; GST;
 KW KM altered culture condition; environmental stress;
 KW KM physiological provocation; ds.
 XX OS Bacillus clausii.
 XX PN WO200229113-A2.
 XX PD 11-APR-2002.
 XX PF 05-OCT-2001; 2001WO-US31437.
 XX PR 06-OCT-2000; 2000US-0680598.
 XX PR 27-MAR-2001; 2001US-279526P.
 XX PA (NOVO) NOVOZYMES BIOTECH INC.
 XX PA (NOVO) NOVOZYMES AS.
 XX PI Berka R, Clausen IG;
 XX DR WPI; 2002-416684/44.
 XX XX
 XX Monitoring differential expression of several genes in first Bacillus cell relative to expression of same genes in one or more second Bacillus cells, by using substrate containing Bacillus genomic sequenced tag array -
 XX Claim 11; SEQ ID NO 4849; 200pp; English.
 XX The invention describes a method of monitoring differential expression of genes in a first Bacillus cell relative to expression of the genes in other Bacillus cells, comprising hybridising labelled nucleic acid probes isolated from Bacillus cells to a substrate containing array of Bacillus genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first Bacillus cell relative to expression of the same genes in one or more second Bacillus cells. The method is useful for monitoring global expression of several genes from a Bacillus cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which Bacillus cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 564 BP; 153 A; 134 C; 141 G; 136 T; 0 other;
 SQ
 Query Match 4.5%; Score 87.6; DB 24; Length 564;
 Best Local Similarity 51.6%; Pred. No. 5.3e-10;
 Matches 263; Conservative 0; Mismatches 229; Indels 18; Gaps 2;
 QY 179 GAGGATATCTTAATGCTAGCGCCTTATCTTCTGGCGCTTGGGCGCTTGACCTGCA 238
 DB |||||
 QY 45 GAGAAATGGCAAGGCGCAAGCGGTGATCTCTCCCTGGGGTGGCGCTTTCTCTGACGCA 104
 DB |||||
 QY 239 ATGGATCTTCTTAATGCTAGCGCCTTATCTTCTGGCGCTTGGGCGCTTGACCTGCA 298
 DB |||||
 QY 105 ATGGATATCTTACACAAACAAAGCTCCAGCCTTTTACACACTGGGTAGCGGAAC 164
 QY 299 CGACCTTCTGGGTATATCTCTGGGATTCAGCTCTTTGAGTCAAGTGAAGAAAT 358
 DB |||||
 QY 165 AAACCAATTACTCGGCATTGCGTGGATGCAACTATTGTTGATCAAGCAGCGGAACAT 224
 DB |||||

QY 359 GGTCCAATTCAGGCTTGGCTTGAATCTCTGGACGGTGGGCGTTTGAATCATCCA-- 416
 |||||
 DB 225 CAACCAACAGCAGGCGCTTGGTTACTTTCCTGGACGGTTCGAACGTTTCTTGGACGACG 284
 |||||
 QY 417 -----ATGGTTTAAGGGTGCCACATATTTGGATGCGATGCCCTTGGATATAAAGGAA 466
 |||||
 DB 285 GCAGAGGGGCACGCCCTACAAAGTGGCGCATATGGGATGAATAAGCTTGAATTCCTATAGG 344
 |||||
 QY 467 GGGTCAGCAATTTAGATGATGTGGGAATCAACATGTGTATTCTTCTCACTCATATCGA 536
 |||||
 DB 345 CCCACACCACTAACAGAGGTGTTCTGAAGGCTAGCTATATTTTGTCTCACTTCTATGTC 404
 |||||
 QY 527 GCCAATGCCGAGGACAAACAAAGAGTGGATTTCATCTACATGTCAGCTATGTCGACGATTTT 586
 |||||
 DB 405 GTCCGTACAG-----AAACGGACATCATCGTTGCAAGCAGTGATTATTACCAAAAGTGC 458
 |||||
 QY 587 ATTGCATCCATTTCAGNAGGGAATGTTTCATGAGTCCCAATTTTCATCCCGAGAGAGTGA 646
 |||||
 DB 459 CCTGCTATTGTGCNAAAAGGCGCAAGTGTATGGCATGCAGTTCATCCCGAAAAAAGCAGC 518
 |||||
 QY 647 GGTGTTGGACTTTCCTCATATTGAGAAGATTT 676
 |||||
 DB 519 ACTGTCGCATGCGCACTGCTGAAACGCTTT 548
 |||||

RESULT 15
 ABN92347
 ID ABN92347 standard; DNA; 777 BP.
 XX AC
 XX ABN92347;
 XX DT 24-JUL-2002 (first entry)
 XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1810.
 XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 XX KW antibacterial; gene therapy; gene; ds.
 XX OS Staphylococcus epidermidis.
 XX PN US6380370-B1.
 XX PD 30-APR-2002.
 XX PF 13-AUG-1998; 98US-0134001.
 XX PR 14-AUG-1997; 97US-055779P.
 XX PR 08-NOV-1997; 97US-064964P.
 XX PA (GENO-) GENOME THERAPEUTICS CORP.
 XX PI Doucette-Stamm LA, Bush D;
 XX DR WPI; 2002-381255/41.
 XX DR P-PSDB; ABP39802.
 XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis polypeptide, useful for diagnosing and treating bacterial infections -
 XX PS Disclosure; SEQ ID 1810; 267pp; English.
 XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The S. epidermidis sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly S. epidermidis infections. The sequences can be used to screen for compounds able to interfere with the S. epidermidis life cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site.

```
XX
SQ Sequence 777 BP; 269 A; 106 C; 173 G; 229 T; 0 other;

Query Match      4.4%; Score 85.6; DB 24; Length 777;
Best Local Similarity 53.2%; Pred. No. 1.6e-09;
Matches 181; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy 1334 TATCAGTGCACGTGTTAATGTTGGACGAGAGAGGGAGACCCATCGAGCTTATGACTAGCT 1393
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1394 AAGGCTGTTGAGGAACCTGGAGCTGGAGAAATATTATTGAACCTGATTCATTGATTGTGATGGT 1453
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1454 CAAGGAAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCT 1513
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1514 GTTATCGCAAGCAGCGGTGCAGGAGTCGCTGATCATTCTCCGAGTCTTTAATGAACCC 1573
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1574 AACGCATCTGCTGCCCTTGCAGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAGGCT 1633
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1634 GTTAAAGAGCATTGTTGAAGGAAGGATTGAAGTTAGAT 1673
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1673 ATTAAGATAAATGCGTGAAGGAGGTATCCTCGTGAGAT 775
```

Search completed: January 20, 2004, 21:50:04
Job time : 552 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 21:18:15 ; Search time 126 Seconds
(without alignments)
6851.948 Million cell updates/sec

Title: US-09-831-233A-1
Perfect score: 1956
Sequence: 1 gcacgaggaactgcaaa.....aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	98.6	5.0	1664976	4	US-08-916-421B-1
C 2	85.6	4.4	777	4	US-09-134-001C-1810
C 3	71.2	3.6	708	4	US-09-252-991A-12223
C 4	68.4	3.5	1830121	4	US-09-557-884-1
C 5	68.4	3.5	1830121	4	US-09-643-990A-1
C 6	66.8	3.4	24417	2	US-08-846-762-1
C 7	63.8	3.3	807	4	US-09-328-352-2061
C 8	63.8	3.3	882	4	US-09-252-991A-6779
C 9	63.8	3.3	936	4	US-09-252-991A-6956
C 10	59	3.0	297	4	US-09-313-294A-6451
C 11	55.4	2.8	1664976	4	US-08-916-421B-1
C 12	54.4	2.8	7218	1	US-08-232-463-14
C 13	53.2	2.7	1559	4	US-09-489-847-42
C 14	51.6	2.6	489	4	US-09-252-991A-12336
C 15	51.6	2.6	798	4	US-09-252-991A-12156
C 16	51	2.6	624	4	US-09-328-352-1928
C 17	50	2.6	473	1	US-08-764-100-16
C 18	50	2.6	4970	1	US-08-764-100-14
C 19	50	2.6	4970	1	US-08-764-100-20
C 20	49.8	2.5	240	1	US-08-628-417-6
C 21	49.8	2.5	6409	4	US-09-967-908A-1
C 22	49.6	2.5	636	3	US-08-998-416-1137
C 23	49.6	2.5	2447	2	US-09-014-969-14
C 24	48.4	2.5	19124	2	US-08-487-826B-13
C 25	48.2	2.5	10640	4	US-09-417-485D-5
C 26	47.4	2.4	1696	4	US-09-835-811-1
C 27	46.8	2.4	2695	4	US-09-706-197-3

C 28	46.6	2.4	1447	4	US-09-443-041A-27	Sequence 27, Appl
C 29	46.4	2.4	1641	1	US-08-300-903A-8	Sequence 8, Appl
C 30	46.4	2.4	1641	4	US-08-988-197-8	Sequence 8, Appl
C 31	46.2	2.4	664	4	US-09-904-615-66	Sequence 66, Appl
C 32	45.4	2.3	5852	1	US-07-867-106-2	Sequence 2, Appl
C 33	45.4	2.3	6243	2	US-09-056-075-1	Sequence 1, Appl
C 34	45.2	2.3	1129	4	US-09-227-357-40	Sequence 40, Appl
C 35	45.2	2.3	2058	2	US-08-749-391-1	Sequence 1, Appl
C 36	45.2	2.3	2058	3	US-09-390-200-1	Sequence 1, Appl
C 37	45	2.3	2773	4	US-09-996-243-178	Sequence 178, Appl
C 38	45	2.3	8920	2	US-08-446-855A-1	Sequence 1, Appl
C 39	45	2.3	8920	3	US-09-150-741-1	Sequence 1, Appl
C 40	44.8	2.3	1117	3	US-09-247-373B-33	Sequence 33, Appl
C 41	44.8	2.3	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 42	44.8	2.3	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 43	44.6	2.3	6671	1	US-08-280-443-1	Sequence 1, Appl
C 44	44.6	2.3	6671	1	US-08-457-459-1	Sequence 1, Appl
C 45	44.6	2.3	6671	1	US-08-555-678-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
;
US-08-916-421B-1

Query Match 5.0%; Score 98.6; DB 4; Length 1664976;
Best Local Similarity 51.0%; Pred. No. 9.8e-14;
Matches 233; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 1259 AGAAAAACCCGATGAAGTGAAGTAAAGCCATCAAGTAAAGCCATCCAGGTCCAAACGGT 1318
Db 1101055 AGACACTATGTTAATGAAGATGAGATAGATAAGATAAATAAATGTCGTTAAAGTAGAG 1100996
Qy 1319 GAGGAATATGCTGGTGTATCAGTCACCTGTTAATGTTGGAGGAGAGAACCCATCCGA 1378
Db 1100995 GATGGTTATTGCTGTTTGAAGTTTATATATACGAGGGAGAGAAACAGGTATAGAT 1100936
Qy 1379 GCTTATCACTAGCTAAGGCTGTTGAGGACTTGGAGCTGAGAAATATTATTGAACCTGC 1438
Db 1100935 GCCATAAACCTGGGCTAAAAAAGTTGAAGAAATTTGGAGCTGGAGAGATTTTATTGCAAGT 1100876
Qy 1439 ATTGATTGTGATGGTCAAGGAAAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGAT 1498
```


Db 315 GCGCTGAAGAGCGGTACGACCTGGCGGTGACCGGCCATCAGCGAGCGGTGAACGTG 256
Qy 1511 CCTGTTATCCGACAGCGGTGACGAGTGGCTGATCACTTCTCCGAAGTCTTTAATGAA 1570
Db 255 CCGGTGATCGTTCGCGCGCGGTGCGCAACCTGGAGCACCTGGCGCGCGCATCCTCGAG 196
Qy 1571 ACCAAGCATCTGCTGCGCTTGACGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAG 1630
Db 195 GCGAAGCGCGGCGGTGCTCGCGCGGAGCATCTTCCACTTGGCGGAGTACACCGTGCCG 136
Qy 1631 GCTGTTAAAGACACTTGTGTAAGAGAGGATTAAGTTAGAT 1673
Db 135 GAAGCAAGGCGTACCTGGCGCGCGGTATCGTGGTGGCT 93

RESULT 9

US-09-252-991A-6956
; Sequence 6956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6956
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6956

Query Match 3.3%; Score 63.8; DB 4; Length 936;

Best Local Similarity 51.6%; Pred. No. 2e-06;
Matches 146; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 1391 GCTAAGGCTGTGAGGAACCTGGAGCTGGAGAAATATTATTGAACCTGATTTGAT 1450
Db 652 GCGAAGAGATGGAAGACCTGGCGCTGGCGAGATTCTCTGACCAAGTGGACCGAC 711
Qy 1451 GGTCAAGAAAGATTCGATATAGATCTGATCAAGCTATATCCGATGCTGTGAACATT 1510
Db 712 GCGGTGAAGAGCGGTACGACCTGGCGGTGACCGGCCATCAGCGAGGCGGTGAACGTG 771
Qy 1511 CCTGTTATCGCAAGCAGCGGTGACGAGTGGCTGATCACTTCTCCGAAGTCTTTAATGAA 1570
Db 772 CCGGTGATCGTTCGCGCGCGGTGCGCAACCTGGAGCACCTGGCGCGCGCATCCTCGAG 831
Qy 1571 ACCAAGCATCTGCTGCGCTTGACGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAG 1630
Db 832 GCGAAGCGCGGCGGTGCTCGCGCGGAGCATCTTCCACTTGGCGGAGTACACCGTGCCG 891
Qy 1631 GCTGTTAAAGACACTTGTGTAAGAGAGGATTAAGTTAGAT 1673
Db 892 GAAGCAAGGCGTACCTGGCGCGCGGTATCGTGGTGGCT 934

RESULT 10

US-09-313-294A-6451
; Sequence 6451, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalngudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313.294A

; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6451
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700351825H1
; NAME/KEY: unsure
; LOCATION: 101, 122
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-6451

Query Match 3.0%; Score 59; DB 4; Length 297;

Best Local Similarity 56.9%; Pred. No. 1.9e-05;
Matches 107; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 29 CCACAGAATGGGTTCAGAATTCGCGCGCTTGGCTGGTGCAGGCGGAGATTCTGTGGTG 88
Db 110 CCGCGATGCGGTGCGGTCTTCCGTCTCGTCCGTGCGTCTGCGCGCAACACGGTG 169
Qy 89 ACTTTACTTGATPACGGTCTCGAAATGTTAGAGTGTGAGGAACCGCCATCCGCACATT 148
Db 170 ACTCTGCTGACTACGCGCGCGGAACGTACGCAGCGTGCACAACTTCGTACCTC 229
Qy 149 GGAATTTGATATCAAGATGTCGAAAGCCAGAGGATTTCTAAATGCTAAGCGCCTTATC 208
Db 230 GGCTTCGATCCGCGAGCGTGCAGAGCCCGGAGGACATCGTCCCGCGGAACCGCTCGTT 289
Qy 209 TTTCTCTGG 216
Db 290 TTTCCCGG 297

RESULT 11

US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature

```
/ LOCATION: (98120)..(98120)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98159)..(98159)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98239)..(98239)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98266)..(98266)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98343)..(98343)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (10398)..(10398)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (148948)..(148948)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (163385)..(163385)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (191989)..(191989)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (191995)..(191995)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (231980)..(231980)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (234187)..(234187)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (234220)..(234220)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (234814)..(234814)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (309398)..(309398)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (309418)..(309418)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (312837)..(312837)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (312993)..(312993)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (319226)..(319226)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (559167)..(559167)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (559241)..(559241)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (600992)..(600992)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (622708)..(622708)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (657081)..(657081)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (657203)..(657203)
```

```
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (674435)..(674435)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (682442)..(682442)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (713652)..(713652)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (741684)..(741684)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (779455)..(779455)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (779676)..(779676)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (855539)..(855539)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (871619)..(871619)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1084830)..(1084830)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1096846)..(1096846)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1119881)..(1119881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1130881)..(1130881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349491)..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1664854)..(1664854)
/ OTHER INFORMATION: n equals a, t, c, or g
/ US-08-916-421B-1
```

Query Match 2.8%; Score 55.4; DB 4; Length 1664976;
Best Local Similarity 50.8%; Pred. No. 0.0044;
Matches 299; Conservative 0; Mismatches 266; Indels 24; Gaps 6;

Qy	93	TACTTGATTAACGGTGTCTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGACACTTGGAT	152
Db	1024660	TAATTGATTAACAACGCGAGGGAATTTGAGAAAGTAT-TCAAAAGGCAGTTGAACCTCTATGAT	1024718
Qy	153	TTGATATCAAGATGTGCAAAAGCCAGAGATATTCCTAAATGCTAAAGCGCTTATCTTTC	212
Db	1024719	AAGGTAATAATAACAACAACAG--TGAGAGATTATTTGGCTTGATGAAGTAATTTCTTAC	1024776
Qy	213	CTGCGGTGGGGCCTTTGACCTTCGAATGATGTTCTTATTCGTAAAGGCTGGCTGAAG	272
Db	1024777	CAGGTGTAGGAATTTTGGTAGTGAATGAAATTTAGTCT-----CATTAAGAGAGA	1024830
Qy	273	CACCTCTGTACTTACATTCAGAATGATCGACCTTTCTGGGTATATGCTGGGATTTGGACG	332
Db	1024831	CAATATACAAAATTTGTTGATGATAGATGTTCCAATCTTAGGATATGTTTAGGAATGCAGA	1024890
Qy	333	TACTCTTTGAGTCAAGTGAAGAAAATGGTCCCAATTCAGGCTCTTGGCTTGATTCCTGGAC	392
Db	1024891	TTTTATTGTAAGAGAGCGAAGAAAAGGAATCAAAAGTTTAGGGATTAATAAAGGCA	1024950
Qy	393	GGGTGGGCGTTTTGAATCATCCAATGGTTTTAAGGGTGCCACATATTTGGATGGCATGCCT	452
Db	1024951	ATGTAATCAAGTTTAA-----GGATGTTGAAAACCTTCCACATATGGGCTGGAATAGT	1025004
Qy	453	TGGATATAAAGGAAGGCTCAGCAATTTTAGATGATGGGGAATCAACAT---GTGTATT	509
Db	1025005	TAAAAATAGTTAAAGATTTGCCACTGTTTGAAGGAATAAAAAACAATAGTTACTTTTACT	1025064
Qy	510	TTGTTCACTCATATCGAGCCAATGCCAGGACAACAAGAGTGGATTTTCATCTACATGCA	569
Db	1025065	TTGTTCATTTCATATCATGTAAATCCAGATG-----AAGATTGTATAGTTGGAAAAACTG	1025118
Qy	570	GCTATGTGTGCGATTTTTATTGTCATCCCAATTCAGAAGGGGAAATGTTTCATGCAGTCCAAATTC	629
Db	1025119	AATATGGAAGAGAGTTCAGGCGTTATAACAAGATATATGCTTTGGCACCAATTC	1025178
Qy	630	ATCCCGAGAGAGTGGAGGTGTTGGACTTTTCCATATTGAGAAGATTTTT	678
Db	1025179	ACCCAGAAAAAAGTGGAAAAAATTTGTTTAAAGATTAAGAAAAATTTTGT	1025227

RESULT 12

```

US-08-232-463-14/C
; Sequence 14, Application US/082322463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, P. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 1559
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-42

Query Match 2.7%; Score 53.2; DB 4; Length 1559;
Best Local Similarity 54.6%; Pred. No. 0.00098;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
QY 1763 TTTATTATTATGTTTGGCTATTGAAATTTATTTATTTATTTATTTTGGCAATTTGTTATTT 1822
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1558 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1499
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1823 GAATAGATTGGATTTTACACCTTGGTGCTGCTGTTTATCTCTAGGCCATGTTTGTG 1882
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1498 TTTTGGCTTCTGATAGATTATTTATTTGTTTGTGCTGATCTGAATTTGGAATTTATA 1439
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1883 GATTATATACAGTGAATTAATAATAAATCGATGAATTTATGCTTTTAAAAAAA 1942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1438 AACATTTTACATGATTATAAATTAATCAAAATTTTAAGAGCTAATTTCTAAAAACAAC 1379
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1943 AAAAAAAAAAAAAA 1956
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1378 ATATCAAAAGCAAA 1365
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14

US-09-252-991A-12336/c
; Sequence 12336, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12336

; LENGTH: 489

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12336

Query Match 2.6%; Score 51.6; DB 4; Length 489;
Best Local Similarity 51.8%; Pred. No. 0.0015;
Matches 117; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```
QY 1391 GCTAAGCGTGTGAGGAACCTGGAGCTGGAGAAATATTATTGAACCTGCATTGTTGAT 1450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 GCGCAAGCGCTCGAAGAGCGTGGAGTGGGTGAGATTTTCTAAATTTCTATTGATCGAGAT 192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1451 GGTCAAGGAAAGATTCGATATAGATCTGATCAAGCTAATATCCGATGCTGGAACATT 1510
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 GGAGTGCAGAAAGGCTTCGACAAAGCTCTAGTGGAAAAATATCGCTTCTAACGTCCTCATGTG 132
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1511 CCTGTTATCCAGCAGCGCTGGAGGAGTCGCTGATCACTTCTCCGAAGTCTTTAATGAA 1570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 CCAGTATCCCTGTGTGGAGCTGGCTCATCGCTGACCTCATCGATCTTTTGTAGCGGT 72
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1571 ACCAAGCATCTGCTGCCCTTGGAGCTGGCAATTTTCCATCGCAAAAG 1616
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 ACGTGTGTGTCGGCAGTAGCGGCGGAGGAGCTATTGCTTTTCCATG 26
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

US-09-252-991A-12156

; Sequence 12156, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12156
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12156

Query Match 2.6%; Score 51.6; DB 4; Length 798;

Best Local Similarity 51.8%; Pred. No. 0.0019;

Matches 117; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```
QY 1391 GCTAAGCGTGTGAGGAACCTGGAGCTGGAGAAATATTATTGAACCTGCATTGTTGAT 1450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 GCGCAAGCGCTCGAAGAGCGTGGAGTGGGTGAGATTTTCTAAATTTCTATTGATCGAGAT 573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1451 GGTCAAGGAAAGATTCGATATAGATCTGATCAAGCTAATATCCGATGCTGGAACATT 1510
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 GGAGTGCAGAAAGGCTTCGACAAAGCTCTAGTGGAAAAATATCGCTTCTAACGTCCTATGTG 633
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1511 CCTGTTATCCGAAGCAGCGGTGCGAGAGTCGCTGATCACTTCTCCGAAGTCTTTAATGAA 1570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 CCAGTATCCCTGTGTGGAGCTGGCTCATCGCTGACTCATCGATCTTTTGTAGCGGT 693
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1571 ACCAAGCATCTGCTGCCCTTGGAGCTGGCAATTTTCCATCGCAAAAG 1616
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 ACGTGTGTGTCGGCAGTAGCGGCGGAGGAGCTATTGCTTTTCCATG 739
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 20, 2004, 23:52:19

Job time : 141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 23:50:02 ; Search time 698 Seconds
(without alignments)

9877.417 Million cell updates/sec

Title: US-09-831-233A-1

Perfect score: 1956

Sequence: 1 gcacgagaagggaactggca.....aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	354	18.1	1506	12	US-10-369-493-46239 Sequence 46239, A
2	335.6	17.2	1656	12	US-10-369-493-25187 Sequence 25187, A
3	303.2	15.5	1732	12	US-10-369-493-27128 Sequence 27128, A
4	102.8	5.3	906	12	US-10-369-493-24907 Sequence 24907, A
5	98.4	5.0	819	12	US-10-369-493-45147 Sequence 45147, A
6	98.2	5.0	759	12	US-10-369-493-47005 Sequence 47005, A
7	95.8	4.9	1030	12	US-10-369-493-36928 Sequence 36928, A
8	92.2	4.7	816	12	US-10-369-493-24674 Sequence 24674, A
9	91.2	4.7	780	12	US-10-369-493-42098 Sequence 42098, A
10	91	4.7	768	12	US-10-369-493-43385 Sequence 43385, A
11	87.6	4.5	564	10	US-09-974-300-4849 Sequence 4849, Ap
12	84.4	4.3	405	10	US-09-974-300-4861 Sequence 4861, Ap
13	84.2	4.3	819	12	US-10-369-493-35008 Sequence 35008, A
14	84	4.3	759	12	US-10-369-493-41126 Sequence 41126, A
15	83.6	4.3	757	10	US-09-974-300-302 Sequence 302, App

16	83.6	4.3	757	10	US-09-974-300-471 Sequence 471, App
17	82.6	4.2	636	10	US-09-974-300-364 Sequence 364, App
18	81.6	4.2	750	12	US-10-369-493-34025 Sequence 34025, A
19	81.4	4.2	753	12	US-10-369-493-40377 Sequence 40377, A
20	77.8	4.0	777	12	US-10-369-493-44423 Sequence 44423, A
21	75	3.8	6346	8	US-08-781-986A-237 Sequence 237, App
22	74.2	3.8	768	12	US-10-369-493-37473 Sequence 37473, A
23	72.4	3.7	768	12	US-10-369-493-31890 Sequence 31890, A
24	70.2	3.6	771	12	US-10-369-493-31035 Sequence 31035, A
25	68.6	3.5	774	12	US-10-369-493-24012 Sequence 24012, A
26	68.6	3.5	780	12	US-10-369-493-33621 Sequence 33621, A
27	68.4	3.5	771	12	US-10-369-493-28278 Sequence 28278, A
28	68.4	3.5	774	10	US-09-738-626-2291 Sequence 2291, Ap
29	68.4	3.5	1830121	15	US-10-329-960-1 Sequence 1, Appli
30	68.4	3.5	3309400	10	US-09-738-626-1 Sequence 1, Appli
31	67	3.4	792	12	US-10-369-493-34476 Sequence 34476, A
32	66.8	3.4	24417	15	US-10-216-209-1 Sequence 1, Appli
33	66.4	3.4	786	12	US-10-369-493-26426 Sequence 26426, A
34	65.4	3.3	640681	10	US-09-790-988-1 Sequence 1, Appli
35	65.2	3.3	750	12	US-10-369-493-34677 Sequence 34677, A
36	65	3.3	765	12	US-10-369-493-43797 Sequence 43797, A
37	65	3.3	783	12	US-10-369-493-40776 Sequence 40776, A
38	64.2	3.3	419	10	US-09-960-352-11234 Sequence 11234, A
39	63.4	3.2	765	12	US-10-369-493-39369 Sequence 39369, A
40	63.4	3.2	765	12	US-10-369-493-39760 Sequence 39760, A
41	63.4	3.2	768	12	US-10-369-493-39003 Sequence 39003, A
42	63.2	3.2	792	12	US-10-369-493-35791 Sequence 35791, A
43	62.8	3.2	768	12	US-10-369-493-31654 Sequence 31654, A
44	62.6	3.2	762	12	US-10-369-493-38811 Sequence 38811, A
45	62.6	3.2	780	12	US-10-369-493-35155 Sequence 35155, A

ALIGNMENTS

RESULT 1

US-10-369-493-46239
; Sequence 46239, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46239
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-46239

Query Match	18.1%	Score	354	DB	12	Length	1506
Best Local Similarity	55.3%	Pred. No.	8.2e-77				
Matches	831	Conservative	0	Mismatches	640	Indels	33
Gaps	6						
QY	198	ACGCCCTTATCTTCTCGGCTTGGCCCTTTGCACCTGCAATGGATGTTCTTATTCGTA	257				
Db	2	AGTGCTGATTTTCTCGGCTTGGAACTTTGGTTTCGTATGCGATTCCTTAGCAAGC	61				
QY	258	AGAGGCTGGCTGAAGCACTCTGTACTTACATTCAGAAATGATCGACCTTTCTGGGTATAT	317				
Db	62	AGGATTTTGAACCTTGGCTGCTATGCTTTGTCTGGCAACCTTTTATGGCTGAT	121				
QY	318	GCGTGGGATTCAGCTACTCTTTGAGTCAAGTGAAGAAATGGTCCAATCAAGGCTTGG	377				

1090 TTTTCAGATCGGCGCCGATTAAGGTTTCGATCGGAAGTGTGACGATTTTACACTGCTCGAGGA 1149
1059 CTTTCAGATCTGGTGTGATTAAGATGATCGATCGGTACGGATGCACTATTCGACGCCGAAA 1118
1150 ATATATTAAACCGGAGTGAAGACA---GGAAGAGCAGCATAGAGCAGATATCTACAGT 1206
1119 ATACTACGAGTTGGGTAAACAGAGAGATGGAACGTCACCAATAGACAAATCTCGAAAGC 1178
1207 ATATGTTAAACCGCAGTGGTGTGAGCATTTGATCTCGCGGATTTTACTTTGAGAAACC 1266
1179 ATACGGTGTACGCGAGTTGTTATTTCTGTCGACCCCTTAAGAGAGTATATGTAATTCACA 1238
1267 CGATGAAGTAAATTTAAAGCCATCAAGTAAGCATTCCAGGTCCAAACCGGTGAGGAATA 1326
1239 AGCAGATACGAAGAACAAAGTCTTCGAGACAGAAATATCCGGGCCCAATCGAGAGAAATA 1298
1327 TGCCTGGTATCAGTCGACCTGTTAATGTTGACAGAGAGGAGACCCATCCGAGCTTATGA 1386
1299 CTGCTGGTACCAATGATCAATCAAGGTGGAAGAGAACTTACGAGACCTTGGTGTGGGA 1358
1387 ACTAGCTTAAGGCTGTGAGGAACCTTGAGCTGGAGAAATATTGAACTGCATTGATTG 1446
1359 ATTAACAAGGGCATGTGAAGCTCTAGGTGCTGGGAGATTTTATGAACTGCTATAGACAA 1418
1447 TGATGGTCAAGGAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGATGCTGTGA 1506
1419 GGATGCTCTAATCTGTTGATGATCTGGAATTTGATAGACATGTTAAAGATCGGTCAA 1478
1507 CATTCCTGTTATCGCAAGCAGCGGTGAGAGTGGTGTGATCACTTCTCCGAAGTCTTTAA 1566
1479 GATTCCTGTCATTCGATCGAGTGGCGCGGTGTACCGGAATTTTCGAAGAGGCTTCT 1538
1567 TGAACCAAGCATCTGCTGCCCTTCAGCTGGCATTTTCCATCGCAAGAGGTTCCAT 1626
1539 AAAGACCCGCGAGATGCTTGGTGGTGAGGTATGTTCCAGAGGTGAATTCATCTGT 1598
1627 TAAGGCTGTTAAAGAGCAGCTGTTTGAAGGAAGGATTTGAAGTTAGATTGTAAGGCGAG 1684
1599 TAACGATGAAGGAGTATTTACTAGAGCAGGATTAAGGTTAGATGATGAAGAG 1656

RESULT 3

US-10-369-493-27128
; Sequence 27128, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27128
; LENGTH: 1732
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-27128

Query Match 15.5%; Score 303.2; DB 12; Length 1732;
Best Local Similarity 58.2%; Pred. No. 3 3e-64;
Matches 552; Conservative 0; Mismatches 393; Indels 3; Gaps 1;
747 GAGTAATGCTTCCTTGAATGAGGCGCAATGATATGGGATCTTGTGTAACCAAG 806
783 GTGTCATTGCTTGTGGATGTGCGCACCAACGACGAGGCGGACTTGGTGTGTCACCAAG 842

807 GAGACCAATATGATGTGAGAGAACGTACAGAAAGAAATGAGGTACAGAAACCTTTGGCAAGC 866
843 GTGACCAATATGATGTCCGGAAGAAAGTCTGACGACCGCAACGTCGCGCAACTCGGCAAGC 902
867 CTGTTGAACCTGCTGGGCAAGTATTTATTAGACGGTCTGTANTGAGGTGAGCTTCTTAAACA 926
903 CCGTCGAGATGGCGCGCAAGTACTACGAAACAGGCGCGGACGAAAGTGACCTTCTCTCAACA 962
927 TTACTGTTTCCGGGACTTCCCTCTAGGCGATCTACCCATGCTACAGGTCTTCGCAAGCG 986
963 TCACCTCTTCCGCGACTGCCCCGTTGCCGACCTGCCCATGCTCGAGATCTTTCGCTCGA 1022
987 CATCTGAAACCGTTTTTGTGCCATTAACCTGTCCGGGGTGGCATCAGGGAATTTTACTGATG 1046
1023 CCTCCAAGACCGTCTTTCGTTCCCTGACCGTCCGGCGCGCATCCGCGACACCGTCGACA 1082
1047 CAAATGGAAGTATTTATTCTAGTCTAGAAAGTGGCTTCAGAGTATTTTCAGATCCGGGCGCG 1106
1083 CGGACGCGCAACAGGTCTCCGCCCTCGAAATCGCCACCATGTACTTCCAGTCAGGCGCG 1142
1107 ATAAAGTTCGATCGGAAGTGTGATGCAAGTTTACACTGCTGAGGAATATATTAAACCGGAG 1166
1143 ACAAGGTGTCCATTTGGCTCCGACCGCTCATCTGCTCCGAGGAGTACTATGCTCTCGGCA 1202
1167 TGAAG---ACAGGAAAGAGCAGCATAGAGCAGATATCTACAGTATATGTTAAACCGAGCAG 1223
1203 AGAGCTCTTTGGCAACACGGCCATTTAGCAGATCAGCAAGCTCAGGNAACCGAGCTG 1262
1224 TGGTTTAAAGATTTGATCTCCGCGAGTTTACTTTGAGAAACCCGATGAAAGTAAATTTA 1283
1263 TCGTCTCTCGGTGATCCAAAGCGGTCTACGTCGCCCAAGCCGATGCCACGAGACACC 1322
1284 AAGCCATCAAGTAAAGCCATCCAGGTCCAAAGCGTGAAGAAATGCTGCTATCAGTGCA 1343
1323 ACACCGTCAAGACGTCGTACCCCGCGCCCAAGGCGGAGGAATACTGCTGTTGTCGCTGCA 1382
1344 CTGTTAATGTTGACGAGAGGAGACCCATCCGAGCTTATGAACTAGCTTAAGGCTGTTG 1403
1383 CGATCAAGGCGGCGCGCAGACGAGGATCTGGATGTTGTTGAATCTCACGAGCGCTGG 1442
1404 AGGAATCTGGAGCTGGAGAAATATTATGAACTGCAATGATTTGATGTTCAAGGAAAAG 1463
1443 AAGCCATGGGCTCGCGCGAGATTTCTGTGAATGCAATGCAAGGACGCGCACCAACAGCG 1502
1464 GATTCGATATAGATCTGATCAAGCTAATATCCGATCTGTGACATTTCTGTTATCGCAA 1523
1503 GGTTCGATCTGGAGTTGATCAGGCAAGTCAAGGCGCGCGGAGGATACCGGTTATTGCCA 1562
1524 GCAGCGGTGACGAGTCTGATCACTTCTCCGAACTCTTTAATGAAACCAACCGCATCTG 1583
1563 GCTCGGGGCGGCAACCCGGGTCACTTTGAGGAGGTGTTAGGAGACGACGACGAGC 1622
1584 CTGCCCCCTGAGCTGGCATTTTCCATTCGCAAGAGAGGTTCCAAATTAAGGCTGTTAAAGAGC 1643
1623 CCGCGCTGGGAGCAGGCATGTTCCACCGTGGGAGTATACAGTACAGCAGGTCAAGGAGG 1682
1644 ACTTGTGAAGGAGGATTTGAATTTAGATTGTAAGCGGAGATCACT 1691
1683 AGCTGAAGCGGAGGGGTTGGTGTGACAGCAGTTTGAGGGAGATCTCT 1730

RESULT 4

US-10-369-493-24907
; Sequence 24907, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 24907

; LENGTH: 906

; TYPE: DNA

; ORGANISM: Methanobacterium thermoautotrophicum

US-10-369-493-24907

Query Match 5.3%; Score 102.8; DB 12; Length 906;

Best Local Similarity 55.2%; Pred. No. 7.7e-15;

Matches 200; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1316 GGTGAGGAATATGCTGGTATCAGTGCACCTGTTAAATGGTGGAGGAGGAGACCCATC 1375

DB 544 GATGATGGATACCTGCTGGTACGAGTGCAGCATCTATGGTGGAGAGATTTCACAGGTATT 603

QY 1376 GAGCTTATGAATAGCTAAGCTGTTGAGGAACCTTGGAGCTGGAGAAATATTATTGAAC 1435

DB 604 GACGCGTGAATGGGCAATGGAAATGCCAGATAGAGTGCAGGTGAATTTCTCTCACA 663

QY 1436 TGCATTGATTGTGATGTCAGGAAAGGATTGATATAGATCTGATCAAGCTAATATCC 1495

DB 664 TCAATGACCGTGAACGATACAAAGATGGGCTACGACATCCCTTCACAGGACCATGAGC 723

QY 1496 GATGCTGTGAACATCTCTGTTATCGCAAGCAGCGGTGCGAGGTGCGTATCACTTCTCC 1555

DB 724 GAGAACCTTGACATACCTGTGATCGCATCAGGGGTGAGGAACTGGAACATCTAT 783

QY 1556 GAAGTCTTTAATGAAACCAACGCATCTGTCGCCCTGACGCTGGCACTTTTCCATCGCAAA 1615

DB 784 GAGCAATTCACAGATGTTAAGGCAGACGCTGCACTTGTCTCAAGCATATTCCACTTCAAT 843

QY 1616 GAGTTCCAAATTAAGCTGTTAAAGACCATCTGTTGAGGAAGGATTGAAGTTAGATTG 1675

DB 844 GAATACCTCTGCTGAGTGAAGGATACCTCAGATCAAGGGGAGTTCCCATAGATTG 903

QY 1676 TA 1677

DB 904 TA 905

RESULT 5

US-10-369-493-45147

; Sequence 45147, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 45147

; LENGTH: 819

; TYPE: DNA

; ORGANISM: Methanococcus jannaschii

US-10-369-493-45147

Query Match

Best Local Similarity 50.8%; Score 98.4; DB 12; Length 819;

Matches 219; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 1259 AGAAACCCGATGAAGTGAATTTAAAGCCATCAAGTAAGCCATCCAGTCCAAACGGT 1318

DB 397 AGACACTATGTTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 456

QY 1319 GAGGAATATGCTGGTATCAGTGCACCTGTTAATGGTGGAGAGAGAGAGAGAGAGAGAG 1378

DB 457 GATGGTTATTGCTGGTTTGAAGTTTATATATACGAGAGGAGAGAGAGAGAGAGAGAG 516

QY 1379 GCTTATGAACCTAGCTAAGCTGTTGAGGAACCTTGGAGCTGGAGAGAGAGAGAGAGAG 1438

DB 517 GCCATAAAGCTGGGCTAAAGAAATTTGAAGAAATTTGGAGCTGGAGAGAGATTTTATTG 576

QY 1439 ATTGATTGTGATGCTCAAGGAAAGGATTCGATATAGATCTGATCAAGCTAATATATCCGAT 1498

DB 577 ATTGATTAAGATGGGCAAAAAGTGGCTATGATTTGATATTGCAAAAGGAGATTTCTAAA 636

QY 1499 GCTGTGAACATCTCTGTTATCGCAAGCAGCGGTGCGAGGATCGCTGATCACTTCTCCGAA 1558

DB 637 AGTGTAAAGCTCCCTGTTATGCAAGTGGAGGTTGTGGAAGAACAGAACATGTTTATGAG 696

QY 1559 GTCTTTAATGAACCAACGCATCTGCTGCCCTTGCAGCTGGCATTTTCCATCCCAAGAG 1618

DB 697 GCATTTTGTATGGAAGGAGATGCCGCATTAATGCGCAGGGATATTGCACTATAGAGAA 756

QY 1619 GTTCCAAATTAAGCTGTTAAAGAGACACTTGTGGAAGGAGGGATTCGAAGTTAGATTGTA 1678

DB 757 TATACANTAGAGAGATAAAAATACTGTGCTGATAGAGGAATACCTATGAGATTATTA 816

RESULT 6

US-10-369-493-47005

; Sequence 47005, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 47005

; LENGTH: 759

; TYPE: DNA

; ORGANISM: Bacillus subtilis

US-10-369-493-47005

Query Match 5.0%; Score 98.2; DB 12; Length 759;

Best Local Similarity 52.8%; Pred. No. 9.5e-14;

Matches 234; Conservative 0; Mismatches 208; Indels 1; Gaps 1;

QY 1232 AGCATTTGATCTCCGCGAGTTTACTTTGAGAA-AACCCGATGAAGTAGAATTTAAAGCCAT 1290

DB 315 AGCTGTTCTCCGCCCTGAGCTGATTACAGAAGGAGCGGAGTTTTTCGTTCTCAATGCAT 374

QY 1291 CAAAGTAAGCATCCAGTCCAAACCGTGGAGGATATGCTGCTATCAGTGCATGTTAA 1350

DB 375 CGTACTTGGCAATTCGATGCCAAGTATGACAAAGAACTGACACATATAAGGTCTACAGCA 434

QY 1351 TGGTGGACGAGAAGGAGACCCATCGAGCTTATGAAGTCTTTCAGGAACT 1410

DB 435 CGCGCGCAGAAAGAAACAGATTGGGAGGTCACCGCGTGGCAAAAGAGCGCTCAACG 494

QY 1411 TGGAGCTGGAGAAATATTATTGAACCTGCATTTGATGTTGATGGTCAAGAGAAAGGATTGCA 1470

DB 495 CGGGCGCAGGAGAAATATTGCTGCAAGCATGGACTCCGACGGTGCAGAAAAAAGGTTTCGA 554

QY 1471 TATAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCTGTTATCGAAGCAGCGG 1530
DB 555 CCACAGCTGACAAAGCTGTTCCGAAGCTGTCGCCGCTGTTATCGTTCGGGAGG 614
QY 1531 TGCAGGAGTCGCTGATCACTTCTCCGAGTCTTTAATGAACCAACGATCTGCTGCCCT 1590
DB 615 CGCGGGAACCGCAGCATATGCTTGAGGCTTTTACAAAAGAGGAGACCCGACGCCGCT 674
QY 1591 TGCAGCTGGCATTTCCATCGCAAGAGGTTCCAAATTAAGGCTTTAAAGAGCACTGTT 1650
DB 675 GGCGCTCAATTTTTCATTAAGAAACATCGAATTAAGAAAGTGAATCATACTTGAA 734
QY 1651 GAAGGAGGATTTGAAGTTAGAT 1673
DB 735 AGATACGGGTGAATGTGAGAT 757

RESULT 7

US-10-369-493-36928
; Sequence 36928, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 36928

; LENGTH: 1030

; TYPE: DNA

; ORGANISM: Aspergillus nidulans

US-10-369-493-36928

Query Match 4.9%; Score 95.8; DB 12; Length 1030;
Best Local Similarity 55.9%; Pred. No. 4.5e-13;
Matches 231; Conservative 0; Mismatches 167; Indels 15; Gaps 2;

QY 578 GAGCATTTTATGATCCATTCAGAGGGAATGTTTCATCGAGTCCAAATTTTCATCCGAG 637
DB 616 GAGAAATTCATCGCGCGATAGCAGAGATAACATCTTTGCGACACAAATTCATCCGAG 675
QY 638 AAGAGTGGAGGTGTTGGACTTTTCCATATTCAGAGAGATTTTGAATGCTGAT-----TCC 691
DB 676 AAGAGCGGCCAAGCAGCGCTACGACCCCTTCGCGCTTTCTTGGACGAGGCTCAGTCCAT 735
QY 692 TTTAAACAAACAAAGACAGAACCAATGAATGGAAGGCTTCTAAACTTGCACAAAGAGATGA 751
DB 736 TCTGTACATTAGAGACTGATTTTGACAGGAGAGAAACGCTCTTACCCGTAGATC 795
QY 752 ATTGCTTGCCTTGATGTGAGGCAATGATGAATGAGGATCTTGTGTAAACAAAGGAGAC 811
DB 796 ATCGCTGCTCTGATGTTCTGATGATGATGTCGCGGATCTCGTTGTGACTAAGGCGCAT 855
QY 812 CAATATGATGTGAGAGACGTACAGAGAGATG-----AGGTCAAGAACCTTGGC 862
DB 856 CAATATGATGTTCCGAGAAAGGATGTCGCGGATGCTGGAGGCAAGTGAAGAACCTTGGGA 915
QY 863 AAGCCTGTTGAACCTGCTGGCCAGTATTTATAGACGGTCTGATGAGGTGAGTCTTCTTA 922
DB 916 AAGCGGTGATATGCTAGAAATATATACAGACAGGGGCGAGATGAGGTGACGTTTAA 975
QY 923 AACATTACTGGTTTCGGGACTTCCCTCTAGGCGCATCTACCCATGCTACAGGT 975
DB 976 AACATCACCTCTTTCAGAAACTGTCCGTTAGCCGACCTCTCCCTATGCTCGAGAT 1028

RESULT 8

US-10-369-493-24674
; Sequence 24674, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 24674

; LENGTH: 816

; TYPE: DNA

; ORGANISM: Archaeoglobus fulgidus

US-10-369-493-24674

Query Match 4.7%; Score 92.2; DB 12; Length 816;

Best Local Similarity 53.8%; Pred. No. 3e-12;

Matches 190; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 1326 ATGCTGTATCAGTCACACTGTTAATGTTGGAGCGAAGGAGAGACCCATCGGAGCTTATG 1385
DB 464 AGCGGTGTTACGAGGTTGTGATTTACGCGGGGAGAAAGCCGTCGCATAGATCGGCTCT 523
QY 1386 AACTAGCTAAGCTGTTGAGAACTTGGAGCTGGAGAAATATTTAAAGCTGCATGATT 1445
DB 524 GGTGGCGGAAGAGGGTTGAGGAGCTTGGCGCTTGGAGAGATTTCTGCTCAGCTCGATGAACA 583
QY 1446 GTGATGTCACAGAAAGGATTCGATATAGATCTGATCAAGCTAATATCGATGCTGTGA 1505
DB 584 GAGCGGACGAAGGATGTTTCGACATCCCNATACGAGAAGATAGCGAGAGGTTA 643
QY 1506 ACATTCCTGTTATCGCAAGCAGCGGTGCGAGGAGTCTGATCACTTCTCCGAAAGTCTTTA 1565
DB 644 ACATTCAGCTCATAGCTTCAGGAGGAGCTGGAACAAAGAGGACACTTCTACGAGGGTTTG 703
QY 1566 ATGAACCAACGATCTGTCGCCCTTGCAGCTGGCATTTTCATCGCAAGAGAGTTCCAA 1625
DB 704 TGAAGGGAAGGCCGATGCTGCTTAGCTGCAAGCATCTTCCACTACAGAGAAATTTGGA 763
QY 1626 TTAAGGCTGTTAAAGAGCACTTGTTCGAAGGAGGATTCGAAGTTAGATTGTA 1678
DB 764 TCAGGGAATTAAGGAGTATCTGGCTGAGAGGGCGCTGCAGGTGAGATTGTGA 816

RESULT 9

US-10-369-493-42098
; Sequence 42098, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

Db 345 CCCACACCACTAACAGAGGTGTTCTCGAAGGCTACGTATATTTTGTCCATTCTCTATGTC 404
Qy 527 GCCAATGCCGAGGACCAACAAGAGTGATTTCTATCATGCGAGCTATGTCGAGATTTT 586
Db 405 GTCGGTACAG-----AAACGGACATCATCGTTGCAAGCAGTGATTTATACCAAGTGTC 458
Qy 587 ATTGCATCCATTTCAGAAAGGAATGTTTCATGCGAGTCCAAATTCATCCCGAGAGAGTGGA 646
Db 459 CTTGCTATTGTGCAAAAAGGCAAGTGTATGGCATGCAAGTTCATCCGCAAAAAGCAGC 518
Qy 647 GGTGTTGGACTTTCATATTTGAGAGATTT 676
Db 519 ACTGTCGGCATGCGACTGCTGAAACGCTTT 548

RESULT 12

US-09-974-300-4861
; Sequence 4861, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4861
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-4861

Query Match 4.3%; Score 84.4; DB 10; Length 405;
Best Local Similarity 54.5%; Pred. No. 1.7e-10;
Matches 169; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy 1350 ATGGTGACGAGAGGAGAGCCCATCGAGCTTATGAACCTAGCTAAGGCTGTTTGAGGAAC 1409
Db 82 ATGGTGCCCAACACAGACTGCTGTTAGGACCGATTTGGGCCAAGGAGGCTGTCCGCT 141
Qy 1410 TTGGAGCTGGAGAAATATTATGAACCTGCAATGATGTTGATGTCAGGAAAGGATTCG 1469
Db 142 TAGGGGCGCGCAACTGCTGTAACCTAGCATGATCAAGATGCGAGCGCAAGCAGGGTTG 201
Qy 1470 ATATAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTTCTGTTATCGCAAGCAGG 1529
Db 202 ATTTGTGTTGTTTAAAGCGCATCAATGAAGTAGTGACAGTGCCCGGTAATTCCTCGTG 261
Qy 1530 GTGCAGAGTCGCTGATCACTTCTCCGAAGCTTTTAAATGAAACCAACGCAATCTGCTGCC 1589
Db 262 GACGAGGCGGCAAGGCACTTTGTTGATGTTTCAACACATATTACCGAGATGCAAGCC 321
Qy 1590 TTGCAGCTGGCAATTTTCCATCGCAAGAGGTTCCAAATTAAGGCTGTTTAAAGACACTTGT 1649
Db 322 TTGGGCGCATGATTTTCCATTACAAAGAACGTCATTTCAAGAAGTAAAGCCCGTTTAC 381
Qy 1650 TGAAGNAGG 1659
Db 382 GGAAGAAGG 391

RESULT 13

US-10-369-493-35008
; Sequence 35008, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35008
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Methanosarcina mazei
US-10-369-493-35008

Query Match 4.3%; Score 84.2; DB 12; Length 819;
Best Local Similarity 53.9%; Pred. No. 2.9e-10;
Matches 173; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
Qy 1327 TGCCTGTTATCATGTCACCTGTTTAATGTTGACGAGGAGAGCCCATCGAGCTTATGA 1386
Db 471 TGCCTGTTATGAGGTTGTAATTTACGAGGCGCAGAGAGGCTACAGGAATCGATCCCGTGCA 530
Qy 1387 ACTAGCTAAGGCTGTTGAGGAAGCTTGGAGCTGGAGAAATATTTAGTCACTGCAATGATTG 1446
Db 531 GTGGGCAAGAAAGCAGAAAGATTGGGTTCCGGAGAAATCTCTCTCACAAGCATGGACCG 590
Qy 1447 TGATGTCGAAGGAAAGGATTTCATATAGATCTGATCAAGCTAATATCCGATCTGTGAA 1506
Db 591 GGATGTCACGCTGTCGCGGTTATGACCTGCCAATTACGAAAAAGCTCTCTGAAGAGCTCGA 650
Qy 1507 CATTCCTGTTATCGCAAGCAGCGGTGCGAGAGTTCGCTGATCACTTCTCGAAAGTCTTTAA 1566
Db 651 TATCCCGATAATAGCTTTCAGGAGGTGTCGAAAAACCTCAGCACATTTACGAAGGATTTTC 710
Qy 1567 TGAACCAACGATCTGTCGCCCTTCAGCTGSCATTTTCCATCGCAAGAGGTTTCCAAT 1626
Db 711 AATTGAAAAGCGATGCTGCACTTCGACCAAGTATCTTCCACTTCAGAGAGTACTCAAT 770
Qy 1627 TAAGGCTGTTAAAGACACTT 1647
Db 771 TAAGGAAGTCAAGAATACTT 791

RESULT 14

US-10-369-493-41126
; Sequence 41126, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 41126
; LENGTH: 759
; TYPE: DNA
; ORGANISM: Bacillus halodurans
US-10-369-493-41126
Query Match 4.3%; Score 84; DB 12; Length 759;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 21:40:56 ; Search time 4247 Seconds
(without alignments)
11193.679 Million cell updates/sec

Title: US-09-831-233A-1
Perfect score: 1956
Sequence: 1 gcacgagaggaactggc...aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_estc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_est3:*
- 12: gb_est4:*
- 13: gb_est5:*
- 14: gb_est6:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_plg:*
- 27: em_gss_vtl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750.4	38.4	2142	11	AY109594 Zea mays
2	498	25.5	780	10	BF274106 GA_Eb002
3	428.4	21.9	638	13	BU046804 PP_LEa002
4	425.6	21.8	668	12	BM878913 P9-G02 Sw

5	335.8	17.2	727	13	BQ112714
6	332.4	17.0	600	12	BG833389
7	330.8	16.9	739	9	AL820092
8	327.2	16.7	533	10	BE340229
9	317.2	16.2	560	10	BF519112
10	311	15.9	729	14	CD453914
11	308.2	15.8	591	12	BM086170
12	301.4	15.4	716	13	BQ869893
13	300.8	15.4	1041	29	CNS06GQF
14	298.4	15.3	569	10	BE353528
15	289	14.8	511	12	BE353725
16	286.8	14.7	434	14	CA925771
17	276	14.1	645	10	BF053343
18	274.2	14.0	562	9	AI995084
19	272.4	13.9	639	13	BQ116797
20	262	13.4	615	9	AW694763
21	254.4	13.0	626	13	BQ578659
22	253.4	13.0	625	13	BQ578541
23	252	12.9	661	9	AU239364
24	251.2	12.8	562	10	BE920620
25	249.2	12.7	460	10	BE999461
26	248.4	12.7	602	13	BQ992987
27	247	12.6	602	13	BU005663
28	245	12.5	570	10	BE449432
29	245	12.5	642	9	AI055415
30	243.8	12.5	595	13	BQ982164
31	240	12.3	589	10	BF050974
32	236.6	12.1	639	9	AW066760
33	235	12.0	386	10	BE523897
34	232	11.9	501	14	H37732
35	230.6	11.8	494	12	BG833391
36	226	11.6	645	10	BE341560
37	224	11.5	598	13	BU971680
38	217.6	11.1	540	9	AJ470077
39	215.8	11.0	557	9	AV546666
40	205.6	10.5	334	10	BG239122
41	201.8	10.3	898	29	CNS06KXT
42	194.4	9.9	544	14	CD448587
43	192.2	9.8	583	13	BQ588079
44	190	9.7	483	10	BF704745
45	189.2	9.7	476	12	BM065703

ALIGNMENTS

RESULT 1	AY109594	AY109594	2142 bp	mrna	linear	HTC 17-OCT-2002
LOCUS	Zea mays	CL287_1	mrna sequence.			
DEFINITION	Zea mays	CL287_1	mrna sequence.			
ACCESSION	AY109594					
VERSION	AY109594.1	GI:21213374				
KEYWORDS	HTC.					
SOURCE	Zea mays					
ORGANISM	Zea mays					
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
AUTHORS	Hayney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsett,M.S.,					
	Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.					
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of					
JOURNAL	Overgo Probes					
REFERENCE	Unpublished (2002)					
AUTHORS	2 (bases 1 to 2142)					
TITLE	Direct Submission					
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of					
COMMENT	Missouri, Columbia, MO 65211, USA					
	If you are interested in getting corresponding physical clones,					
	these are publicly available from ZmDB and may be found by BLAST					
	searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,					
	www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the					

RESULT 2
BF274106
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

BF274106 780 bp mRNA linear EST 07-MAR-2001
GA_Eb0020C04f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Eb0020C04f, mRNA sequence.
BF274106
BF274106.1 GI:11205176
EST.
Gossypium arboreum

ORGANISM
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.

REFERENCE
1 (bases 1 to 780)

AUTHORS
Wang,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry D., Wood,T.C., Leslie,A. and Wilkins,F.A.

TITLE
An integrated analysis of the genetics, development, and evolution of the cotton fiber

JOURNAL
Unpublished

COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCATATAGGG
High quality sequence stop: 726.

FEATURES
source
1..780
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Eb0020C04f"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/notes="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI"

BASE COUNT
231 a 123 c 206 g 220 t

ORIGIN
Query Match 25.5%; Score 498; DB 10; Length 780;
Best Local Similarity 79.2%; Pred. No. 1.3e-51;
Matches 616; Conservative 0; Mismatches 160; Indels 2; Gaps 2;

QY 881 GGACAGTATTATTAGACGGTGTGATGAGTGTGAGTCTTAAACATTACTGTTCCGG 940
DB 1 GGACAGTATTACAAAGTGGGGCGATGAGTGTGAGTCTTAAACATTACTGTTCCGGT 60

QY 941 GACTTCCTCTAGCGCATCTACCATCTACAGTCTTTCGACGCGCATCTGAAACGTT 1000
DB 61 GATTTCCTCTAGCGCATCTACCATCTACAGTCTTTCGACGCGCATCTGAAACGTT 120

QY 1001 TTGTGCGCATTAATCTGCGGGGTGGCATCAGGATTTTACTGATGCAAAATGAAAGTAT 1060
DB 121 TTGTGCGCGTAAACAGTTGGAGGTGGTATACGGGATTTTACTGATGCAAAATGCGGTAC 180

QY 1061 TATCTAGTCTAGAGTGGCTTCAGAGTATTTTCAGATCGGGCGCGCATAGTTCGATC 1120
DB 181 TATCTAGTCTAGAGTGGCTTCAGAGTATTTTTCAGATCGGGCGCGCATAGTTCGATC 240

QY 1121 GGAAGTGTAGTCACTGTGAGGATATATTAACACCGAGTGAAGACAGGAAAG 1180
DB 241 GGAAGTGTAGTCACTGTGAGGATATATTAACACCGAGTGAAGACAGGAAAG 300

QY 1181 AGCAGCATAGCAGGATATCTACAGTATATGTTAAACAGCAGTGGTGTGAAGCATGAT 1240
DB 301 AGCAGCTTAGACAAATTTCTAAAGTTTATGGAATTCAGGAGTGGTGTGAAGCATGAT 360

QY 1241 CCTCGCGGATTTACTTGAGAAACCGATGAGTATTTAAGCCATCAAGTAAGC 1300
DB 361 CCTCGTAGGCTGATGTTCAAGGTCTTAATGATGTCAGTTCAGGACCATACCGGTCCTCA 420

QY 1301 CATCCAGTCCAAACCGTGAAGGATATGCTGGTATCAGTGCATGTTAATGTTGACGGA 1360
DB 421 AAACAGTCCAAATCGAGAGATATGCTGGTATCAGTGTACGTTAATGTTGACGGA 480

QY 1361 GAAGGGAGACCCATCGAGCTTATGAACACTAGCTAAGCTGTGAGGAACTTGGAGCTGGA 1420

Db 481 GAAGGGCGATCAATTGGAGCTTACGAGCTTGGCGAAAGTTGTTGAAGAACTGGAGCTGGA 540

QY 1421 GAAATATTATTGAATGCTGATTTGATGATGTTGATGTTGATGTTGATGATGATGATGATG 1480

Db 541 GAAATATTATTGAATGCTGATTTGATGATGTTGATGATGTTGATGATGATGATGATGATG 600

QY 1481 ATCAAGCTAATTCGATGCTGTGAACATTCCTGTTATCGAAGACGCGGTGAGGATC 1540

Db 601 ATAAAGCTATATCTGATGCTGTCAACA-TCCTTGTATTGCAAGTAGTACGCTGCTGCTGCT 659

QY 1541 GCTGATCACTTCTCGAAGTCTTTATGAACCAACGCAATCTGCTGCCCTTCAGCTGGC 1600

Db 660 GTTGAACACTTCTCAGAGGT-ATTATGAACCAATGCAATGCTGCTGCTTGTGCTGCTGCT 718

QY 1601 ATTTTCCATCGCAAGAGGTTTCCAATTAAAGGCTGTTAAAGAGCACTTGTGTAAGGAAG 1658

Db 719 ATTTTCCATCGGAGGAGTGGCCATTCAGTCTGTAAGAGCACACTTGTGTAAGGAAG 776

RESULT 3
BU046804
LOCUS
PP_LEA0027M01f Peach developing fruit mesocarp Prunus persica cDNA
DEFINITION
BU046804
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Prunus persica (peach)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
1 (bases 1 to 638)
Callahan,A., Palmer,M., Main,D., Wing,R. and Abbott,A.
Peach Model Genome for Rosaceae
Unpublished
Contact: Abbott, A.
Dept of Genetics and Biochemistry
Clemson University
122 Long Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 3060
Fax: 864 656 6879
Email: aalbert@clemson.edu
Total High Quality bases = 513
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 638.

FEATURES
source
1..638
/organism="Prunus persica"
/mol_type="mRNA"
/cultivar="Loring"
/db_xref="taxon:3760"
/clone="PP_LEA0027M01f"
/tissue_type="Mesocarp"
/lab_host="E. coli"
/clone_lib="Peach developing fruit mesocarp"
/note="Vector: pBluescript II SK(-); Site 1: EcoRI; Site 2: XhoI; authority=Prunus persica L. Bateh; The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis go to
http://www.genome.clemson.edu/projects/peach. To order this clone go to http://www.genome.clemson.edu/orders"

BASE COUNT
188 a 92 c 174 g 182 t

ORIGIN
Query Match 21.9%; Score 428.4; DB 13; Length 638;
Best Local Similarity 79.7%; Pred. No. 4.3e-43;
Matches 504; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 905 GATGAGGTGAGTCTTAAACATTACTGTTTCGGGACTTCCCTTAGCGGATTCACCC 964

Db 1 GATGAGGTGAGTGTCTTTTAAATATTACCGGTTTCGCGACTTCCCTTTGGCGGACTTGCCA 60
Qy 965 ATGCTACAGGTCTTGCACCGGCATCTGAAACGTTTTTGTGCCATTAACTGTCGGGGT 1024
Db 61 ATGTTACAGGTACTGAGATACATCAGAAATGTTTTGTACCAATTAACAGTGGAGGT 120
Qy 1025 GGCATCAGGATTTTACTGATGCAATGGAAGGTATTATTCTAGTCTAGAGTGGCTTCA 1084
Db 121 GGCATTAGAGATTTTACAGATGCTAATGCGAGAGTATTCTAGTTGGAGTGGCTTCA 180
Qy 1085 GAGTATTTCAGATCGGGCGCGGATGAGTTTCGATCGGAAGTATGATGAGTTTACACTGCT 1144
Db 181 GAATATTTCAGATGTCGGGCTGATAAGATTCCATTGGGAGTGATGAGTTTATGCTGCA 240
Qy 1145 GAGGAATATTAAACCGGAGTGAACACAGAGAAAGACATAGAGCAGATATCTACA 1204
Db 241 GAAGAAATATTAAAGACTGGAGTAAAGATGGAATATAGTCTTAGAGCAGATATCTAGA 300
Qy 1205 GTATATGTAACACGAGGAGTGTGTTAAGCATTCATCTCCGAGTTTACTTTGAGAAAA 1264
Db 301 GTTTATGGAATACAGCTGTGGTTGTTAAGCATTCATCTCCGAGTGTACCTTAAAT 360
Qy 1265 CCCGATGAAGTGAATTTAAAGCCATCAAGTAAAGCCATCCAGTCCAAACGAGGAA 1324
Db 361 CCAGAGATGTAGGTTCAAGACTATTAGGTTAAACAAACCCAGTCCAAACGAGGAA 420
Qy 1325 TATGCTGTTATCAGTCACTGTTTAAATGTTGGAGAGAGAGGAGCCATCGGAGCTTAT 1384
Db 421 TTTGATGTTATCAGTGTACAGTTAGCGGTGGCGAGAGGCGGACCAATTTGAGGCTTAT 480
Qy 1385 GAATAGTAAAGGCTGTTGAGGACTTTGAGCTGGAGAAATATTATTGAACCTGCATTGAT 1444
Db 481 GAGCTTGAAGAGAGCTTTGAGGACTTTGAGCTGGAGAAATATTGCTTAAACTGCATTGAT 540
Qy 1445 TGTGATGTTCAAGGAAAGGATTCGATATAGATCTGATCAAGCTTAATATCCGATGCTGTG 1504
Db 541 TGTGATGTTCAAGGAAAGGATTTGATATAGATTTAAAGCTGATCTCAGATGCTGTG 600
Qy 1505 AACATGCTGTTATCGAAGCAGCGGTGCGAG 1536
Db 601 AGCATTCCTGTGATTGCTAGTGTGCTGG 632

RESULT 4
BM878913 668 bp mRNA linear EST 07-MAR-2002
LOCUS P9-G02 Sweetpotato Ipomoea batatas cDNA similar to Glutamine
DEFINITION amidotransferase/cyclase, mRNA sequence.
BM878913

ACCESSION BM878913.1 GI:19246579
VERSION EST.
KEYWORDS Ipomoea batatas (sweet potato)
SOURCE Ipomoea batatas

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Convolvulaceae; Ipomoea.

REFERENCE Jarret,R.L., Dang,P., Seo,Y.W. and Jang,C.S.
AUTHORS Sweetpotato ESTs
JOURNAL Unpublished
COMMENT Contact: Jarret RL
Bldg 4457

USDA/ARS, Plant Genetic Resources
1109 Experiment Street, Griffin, GA 30223, USA
Tel: 770/228-7303
Fax: 770/228-7282
Email: bjarret@aes.griffin.peachnet.edu.

FEATURES
source 1..668
/organism="Ipomoea batatas"
/mol_type="mRNA"
/cultivar="Jewel"
/db_xref="taxon:4120"

BASE COUNT 193 a 100 c 186 g 189 t
ORIGIN
Query Match 21.8%; Score 425.6; DB 12; Length 668;
Best Local Similarity 78.3%; Pred. No. 9.1e-43;
Matches 523; Conservative 0; Mismatches 144; Indels 1; Gaps 1;
Qy 886 GTATTATTTAGACGGTCTGATGAGTTCAG-CTTCTTAAACATTACTGTTTCCGGGACT 944
Db 1 GTATTACAAGGATGGAGCTGATGAGGTTAGCCCTTTTGAATATTACTGTTTCAGAGACT 60
Qy 945 TCCTCTACGCGATCTTACCCATGCTTACAGGTCTTTCACACGCGCATCTGAAACGTTTCTTG 1004
Db 61 TCCTCTCGCGCATCTTACCAATGTTCCAGTATTAGGTTACATCAGAGATGTTTCTTG 120
Qy 1005 TGCCATTAACTGTGCGGGGTGGCATCAGGGATTTTACTGATCAATGGAAGTATTATT 1064
Db 121 TGCCTTTAACTGTTGGAGCGGTATTAGAGACTTCACTGATGGAATGCGAGTACTACT 180
Qy 1065 CTAGTCTAGAAGTGGCTTCAGAGTATTTCAGATCGGGCCCGATAGGTTTCGATCGGAA 1124
Db 181 CTAGTTTGAAGTGTCTTCTGAGTATTTTCAGATCTGCTGCTGACAAAGGTCTCTATAGGAA 240
Qy 1125 GTGATGCACTTTACACTGCTGAGGAATATATTAAACCGGAGTGAAGACAGGAAAGAGCA 1184
Db 241 GTGATGCACTATATGCTGCTGAAGATATTGGAATCTGGAGTAAACCTGGNAAGAGCA 300
Qy 1185 GCATAGAGCAGATATCTACAGTATATGTTAAACAGGACGTGTTGTAAGCATTTGATCTTC 1244
Db 301 GTCTGGAGCAGATCTCTAGTGTATTGGAATCAGGACGTGTTGTAAGTATTGATCCAA 360
Qy 1245 GCGAGTCTTACTTGAGAAACCCGATGAAGTAGAATTTAAAGCCATCAAGTAAGCCATC 1304
Db 361 GGAGGGTGTATTGGAAGGCCCAAGAGATGTAGAGTTCAAGACTGTCAAAGTGCAGAACCC 420
Qy 1305 CAGGTCCTCAACGCTGAGGAATATGCTGCTATCAGTGCACCTGTTAATGTTGGACGAGAG 1364
Db 421 TAGTCCAAATGCTGAAGATATGCTTGGTACAGTGCACAGTGAATGTTGGACGAGAG 480
Qy 1365 GGAGACCCATCGGAGCTTATGAACCTAGCTTAAAGCTGTTGAGGAACCTTGAGAGTGGAGAAA 1424
Db 481 GGAGACCAATTTGGAGCATTTGAGCTTGCAGAAAGCTGTTGAAGAAATTTGGGAGCTGGTGA 540
Qy 1425 TATTATTGAACCTGATTTGATGCTCAAGGAAAGGATTCGATATAGATCTGATCA 1484
Db 541 TACTGTAAACCTGCATCGACTCGGATGGTCAAGGTAAGGATTTGATATAGACCTTATCA 600
Qy 1485 AGCTAATATCCGATGCTGTGAACATTCCTGTTATCGCAAGCAGCGGTGCGAGGAGTCGCTG 1544
Db 601 AGTTAATATCAGATGCTGTAAGCATCCAGTAATTCGAAGCAGTGGTGTGGAGCTGTTG 660
Qy 1545 ATCACTTC 1552
Db 661 AGCATTTTC 668

RESULT 5
BQ112714 727 bp mRNA linear EST 07-MAR-2003
LOCUS EST598290 mixed potato tissues Solanum tuberosum cDNA clone STMCI95
DEFINITION 5' end, mRNA sequence.

ACCESSION BQ112714
VERSION BQ112714.2 GI:21915593
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)

ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 727)
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,


```

QY 1337 CAGTGCACTGTTAATGTGGACGAGAGGAGACCCATCGAGCTTATGAAGCTAGCTAAG 1396
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
359 CAGTGCACTGTTAATGTGGACGAGAGGAGACCCATCGAGCTTATGAAGCTAGCTAAG 300
QY 1397 GCTGTTGAGGAACCTTGGAGCTGGAGAAATATTATTGAAGCTGCAATGTTGATGTTCAA 1456
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 GCTGTTGAGGAACCTTGGAGCTGGAGAAATATTATTGAAGCTGCAATGTTGATGTTCAA 240
QY 1457 GGAAGAGATTCGATATAGATCTGATCAAGCTAATATCCGATGCTGGAACATTCCTGTT 1516
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
239 GGTGTTGGAATTCGATATAGATTTGGTTAAATGGTTTCTGATGCTGGAACATCCCTGTC 180
QY 1517 ATCGCAAGCAGCGTGCAGAGTGCCTGATCACTTCTCCGAAGCTTTAATGAACCAAC 1576
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
179 ATTGCGAGCAGTGGTCTGGAGCTGTTCAACATTTTCTGAAATTTTGGAGAAACAAAT 120
QY 1577 GCATCTGCTGCTTGCCTTGCAGCTGGCAATTTCCATCGCAAGAGGTTCCAAATTAAGGCTGTT 1636
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 GCTTCTGCTGCTTGTCTGCTGGCAATTTCCACCGGAAGAGGTTCTTACTAGCAGTG 60
QY 1637 AAAGAGCATTGTTGAAGGAGGATTTGAAGTTAGATTGAAGCGGAGAGATCACTGG 1694
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 AAAGAGCATTGTTGAAGGAGGATTTGAAGTTAGATTGAAGCGGAGAGATCACTGG 2

RESULT 7
AL820092 739 bp mRNA linear EST 15-JUL-2002
LOCUS AL820092 N:130 Triticum aestivum cDNA clone F05_N130_plate_66, mRNA
DEFINITION
ACCESSION AL820092
VERSION AL820092.1 GI:21831692
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE 1 (bases 1 to 739)
AUTHORS Wilson,I., Beswick,R., Shepherd,S., Barker,G., Parker,J., Owen,P.,
Edwards,D., Coghill,J., Holdsworth,M., Lenton,J., Shewry,P. and
Edwards,K.
TITLE A BSRP-funded wheat EST resource for the academic community
JOURNAL Unpublished
COMMENT Contact: Barker G
Institute of Arable Crop Research
Long Ashton, Bristol BS41 9AF United Kingdom.
FEATURES
Source 1..739
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="mercia"
/db_xref="taxon:4565"
/clone="F05_N130_plate_66"
/tissue_type="embryo"
/dev_stage="2 days post germination"
/clone_lib="N:130"
BASE COUNT 206 a 113 c 208 g 212 t
ORIGIN
Query Match 16.9%; Score 330.8; DB 9; Length 739;
Best Local Similarity 75.1%; Pred. No. 2.9e-31;
Matches 452; Conservative 0; Mismatches 147; Indels 3; Gaps 3;

QY 1075 AGTGGCTTCAGAGTATTTCAGATCGGGCGCGGATAGGTTTCGATCGGAAGTGAATGCA 1134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 AGTAGCATCAGAAATATTTCAGTCTGGTCTGGTGAAGATTTCAAT-GGAAGTGAATGCTGT 60
QY 1135 TTACACTGCTGAGGAATATATTAAACCGGAGTGTAAGACAGGAGGAGGAGGAGGAGGAGG 1194
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 TT-TGCTGCTGAAGCCTATTTCAGAGCTGGTGTGAAGACAGGAGGAGGAGGAGGAGGAGGAG 119
QY 1195 GATATCTACAGTATATGTAACCGAGGAGTGGTTGTAAGCATTTGATCTCGCCGAGTTTA 1254

```

```

Db 120 AATTTCTAGAGTATATGGAATCAAGCTCTAGTTTGAAGTATTGATCTCGAAGAGTATA 179
QY 1255 CTTGAGAAAACCGAGTAGAATTTAAAGCCATCAAGTAAGCCATCCAGGTCCTCAAA 1314
Db 180 TGTAAAAGTCCAGATGAGTGCATTTTAAAACGTGTAATTTGTCCAGTAAAAGTCCATC 239
QY 1315 CGGTGAGGAATATGCTGCTGATCAGTGCACTGTTAATGCTGGACAGAGAGGAGAGCCAT 1374
Db 240 AGAGAGAAGATATGCAATGTTACCAAGTGCAAGTAAGTGGTGGCGGTGATAGCGGCTAT 299
QY 1375 CGAGGCTTATGAACCTAGCTTAAGGCTGTTGAGGAACCTTGAAGCTGGAGAAATATTATGA 1434
Db 300 AGGAGCATATGAACCTAGCAAGAGCTGTCGAAGAAATTTGGCGGAGGAGAAATACTACTCA 359
QY 1435 CTGCATTGATTGATGTTGATGTTCAAGGAAGAGATTCGATATAGATCTGATCAAGCTAATATC 1494
Db 360 CTGCATTGATTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTG 419
QY 1495 CGATGCTGTGAACATTTCTGTTATCGCAAGCAGCGTGCAGGAGTGCCTGATCA-CTTCT 1553
Db 420 GGATGCTGTGACAAATTTCTGTTATTCGAAGCAGTGGTCTGCTGCTGCTGCTGCTGCTGCT 479
QY 1554 CCGAAGCTCTTTAATGAACCAAGCAGCTCTGCTGCCCTTGCAGCTGGCAATTTTCCATCGCA 1613
Db 480 CTGAAGTCTTTGAGAAAACAAATGCTTCTGCTGCCCTTGCCTGCTGCTGCTGCTGCTGCTG 539
QY 1614 AGAGGTTCCAAATTAAGGCTGTTAAAGACACTTGTGAAGGAGGATGAAGTTAGAT 1673
Db 540 AAAGGTTCCCAATTTCTAGCAGTGAAGAGCATCTGTTAGTCTGTTGTTGTTGTTGTTGTTG 599
QY 1674 TG 1675
Db 600 TG 601

RESULT 8
BE340229 533 bp mRNA linear EST 10-MAR-2003
LOCUS BE340229 potato stolon, Cornell University Solanum tuberosum cDNA
DEFINITION clone cSTAL0F11 similar to glutamine amidotransferase/cyclase
{Arabidopsis thaliana} GP|7269543|emb|CAB79545, mRNA sequence.
ACCESSION BE340229
VERSION BE340229.1 GI:9249760
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
REFERENCE 1 (bases 1 to 533)
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Holt,I.B., Liang,F., Hansen,T.S., Utterback,F., Bowman,C.L., Doan
B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
B.
TITLE Generation of ESTs from potato swelling stolons
JOURNAL Unpublished
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/.
FEATURES
Location/Qualifiers
1..533
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTAL0F11"
/tissue_type="axillary buds of stem explants, swelling
stolons"

```

/dev_stage="1 to 3 days"
 /lab_host="SOLR"
 /clone_lib="potato stolon, Cornell University"
 /notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; RNA was supplied by Christian Bachem & Beatriz
 Horvath(Laboratory of Plant Breeding, Dept. of Plant
 Sciences, Wageningen University, The Netherlands). Total
 RNA was isolated from developing axillary buds of potato
 nodal stem cuttings cultured on medium for the
 introduction of tuber formation as described in Bachem et
 al. (Plant Journal 1996). Tissue samples were taken of
 stages corresponding to growing stolons and the early
 stages of tuber formation."
 147 a 88 c 154 g 144 t

BASE COUNT 147 a 88 c 154 g 144 t
 ORIGIN
 Query Match 16.7%; Score 327.2; DB 10; Length 533;
 Best Local Similarity 75.9%; Pred. No. 9.6e-31;
 Matches 404; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
 QY 854 AACCTTGGCAAGCCTGTGAACCTTGGCGCAGTATTATTAGACGGTGCTGATGAGTTC 913
 DB 2 AACCTTGGCAAGCCTGTGAACCTTGGCGCAGTATTATTAGACGGTGCTGATGAGTTC 61
 QY 914 AGCTTCTTAACATTAAGTCTTCCGGACCTTCCCTCTAGCGCATCTACCCATGCTACAG 973
 DB 62 AGCTTCTTAACATTAAGTCTTCCGGACCTTCCCTCTAGCGCATCTACCCATGCTACAG 121
 QY 974 GTCTTCAACGCCCATCTGAACAGTCTTGTGCCATTAACCTGCGGGTGCCATCAGG 1033
 DB 122 GTATTGAGGTATGATCAGAGAAATGTTTTGTGCCATTAACCTGCGGGTGCCATCAGG 181
 QY 1034 GATTTTACTGATCAATGAAGGTATTTCTAGTCTAGAGTGGCTTCAGAGTATTTTC 1093
 DB 182 GACTTCAAGATGCGAATGCGATACTACTCTAGTTTGAAGTGTCTTCAGAAATTTT 241
 QY 1094 AGATCGGGCGCCGATGAAGTTCGATCGGAAGTGATGATCTTTACACTGCTGAGGAATAT 1153
 DB 242 CGTTCTGGGCGACACAAAATTTCTATTGGAAGTGATGCTGTTTATGCTGGGAAGATAC 301
 QY 1154 ATTAAACCGGAGTGAAGACGAAAGACGAGCATAGACGAGATATCTACAGTATATGTT 1213
 DB 302 TTAATAATCTGGAGTGAACCTGGAAGACGAGCTAGACGAGATCTCCGAGTTTATGGA 361
 QY 1214 AACCGGCGAGTGTGTGAACATTTGATCTCGCGAGTTTACTTTGAGAAACCCGATGAA 1273
 DB 362 ATACGCGAGTGTGTGAACATTTGATCTCGCGAGTTTACTTTGAGAAACCCGATGAA 421
 QY 1274 GTAGAAATTTAAAGCCCATCAAGTAAGCCATCCAGGTCCAAACCGTGAGGAATATGCTGG 1333
 DB 422 GTGGAATTCAGGACTGTGAAGTGAGAAACCCAGGTCCAAATCGAGAGATGATGATGG 481
 QY 1334 TATCAGTGCACTGTTAATGTGACGACGAGAGGAGACCCATCGAGGTTATG 1385
 DB 482 TACCAGTGCAGGTGATGTTGGGAGAGAGGCGGCCCTATTGAGGCTTATG 533

RESULT 9
 BFS19112
 LOCUS EST456572 DSIL Medicago truncatula cDNA clone pBSIL-19B12, mRNA
 DEFINITION sequence.
 ACCESSION BFS19112
 VERSION BFS19112.1 GI:11607795
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 560)
 AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Peng

H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
 Holt, I.B. and Fraser, C.M.
 ESTs from leaves of Medicago truncatula after inoculation with
 Colletotrichum trifolii
 Unpublished
 Contact: Deborah A. Samac
 Department of Plant Pathology
 University of Minnesota
 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
 Tel: 612 625 1243
 Fax: 651 649 5058
 Email: debysopuccini.crl.umn.edu
 University of Minnesota name: M276541e
 TIGR sequence name: MTFCD06TK
 More information is available at: http://chrysie.tamu.edu/medicago
 Seq primer: SKmod (CTA gAA CTA gtg gat CC).

FEATURES
 source
 1..560
 Location/Qualifiers
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pDSIL-19B12"
 /tissue_type="leaves infected with Colletotrichum trifolii"
 /dev_stage="cotyledons and primary leaves harvested 5 and 8 days after inoculation with Colletotrichum trifolii"
 /lab_host="E. coli strain XL0LR"
 /clone_lib="DSIL"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA from cotyledons and primary leaves harvested 5 and 8 days after inoculation with Colletotrichum trifolii. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using EX-Assist helper phage and propagated in XL0LR cells. Note: EST may be of fungal origin."
 170 a 96 c 129 g 165 t

Query Match 16.2%; Score 317.2; DB 10; Length 560;
 Best Local Similarity 76.7%; Pred. No. 1.5e-29;
 Matches 388; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 QY 1173 CAGGAAAGACGACATAGACGAGATATCTACAGTATATGTTACACGAGCAGTGTGTTAA 1232
 DB 1 CAGGAAAGACGACATAGACGAGATATCAATAGTTTATGGGAATCAGGCGAGTGTGTTA 60
 QY 1233 GCATTGATCCTCGCGAGTTTACTTGAGAAACCCGATGAAGTAGAATTTAAAGCCCATCA 1292
 DB 61 GTATTGATCCCCGTATGGTTTACATAAACCAATGATGTGCAATTTCCAGACCATTA 120
 QY 1293 AAGTAAGCCATCCAGGTCCAAACCGTGAGGAATATGCTCTGGTATCAGTCACTGTTAAATG 1352
 DB 121 AGTTTCAAGTCCAGGTCCAAATGGAGAAATATGCTGGTATCAGTACAGTTAAATG 180
 QY 1353 GTGGACGAGAGGAGACCCATCGGAGCTTATGAACCTAGCTAGGCTGTTGAGGAACCTTG 1412
 DB 181 GGGGGCGAGAGGTCGGCAATTTGGTCTTATGAACCTGGCGAAAGCTGTTGAAGAACTTG 240
 QY 1413 GAGCTGGAGAAATATTATTGAACCTGATGTTGATGCTCAAGGAAAGGATTCGATA 1472
 DB 241 GTGCTGGTGAATACTATTAAATGCAATGATGTGATGTTCAAGGAAAGGATTCATG 300
 QY 1473 TAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCTCTGTATCGCAAGCAGCGGTG 1532
 DB 301 TAGCTTTAATTAAATGATCTCAATCGGCTCAATATCCCTGTGATGCAAGTAGTGGTG 360
 QY 1533 CAGGAGTCGCTCATCACTTCTCCGAAGTCTTTTAATGAACCAACGCAATCTGTCGCCCTTG 1592
 DB 361 CTGGCATTCCTAATCACTTTTCCGATGTGTTCACTAAACCAATGATCATCGCTGCACCTTG 420

```

1593 CAGCTGCGATTTTCCATCGCAAGAGGTTCCAAATTAAGGCTGTATAAGAGCACTTGTGTA 1652
      |||
421 CTCTGCGATTTTTCATACGAGAGGAGTACTATTCTACTCAGTGAAGAACACTTTGTGTA 480
      |||
1653 AGGAAGGGATTGAAGTTAGATTGTA 1678
      |||
481 ATGAGGCATAGAGTCCGAATCTGA 506
      |||

RESULT 10
CD453914/c
LOCUS
DEFINITION
WHE0904_B09_D18ZT CS wheat 5-15 DAP spike cDNA library Triticum
aestivum cDNA clone WHE0904_B09_D18, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
CD453914.1 GI:31368542
EST.
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 729)
Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Crossman
C., Fenton, R.D., Lazo, G.R., Pham, J., Raush, C.J. and Woo, J.
The structure and function of the expressed portion of the wheat
genomes - 5-15 DAP CS spike cDNA library
Unpublished
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
This EST was generated by sequencing from the 3' end of the clone.
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20.
Seq primer: T7 primer.
FEATURES
    source
        1..729
            Location/Qualifiers
                /organism="Triticum aestivum"
                /mol_type="mRNA"
                /cultivar="Chinese Spring"
                /db_xref="taxon:4565"
                /clones="WHE0904_B09_D18"
                /tissue_type="Spike"
                /dev_stage="Adult plant"
                /lab_host="E. coli SOLR"
                /clone_lib="CS wheat 5-15 DAP spike cDNA library"
                /notes="vector: lambda Uni-ZAP XR, excised phagemid;
                Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
                greenhouse. Spikes at 5, 10 and 15 DAP were harvested,
                total RNA and poly(A) RNA were prepared, a cDNA library
                was made, and the cDNA clones were in vivo excised to give
                phagescript phagemids in the TJ Close lab (Choi, Close,
                Fenton) at the University of California, Riverside.
                Plasmid DNA preparations and DNA sequencing were performed
                in the OD Anderson lab (all other authors)."
```

BASE COUNT 199 a 196 c 109 g 225 t

Query Match 15.9%; Score 311; DB 14; Length 729;
 Best Local Similarity 74.4%; Pred. No. 7.3e-29;
 Matches 392; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```

1151 TATATTAAACCGAGTTCAGACAGGAAAGCAGCATAGCAGATCTCAGATATAT 1210
      |||
726 TATTACAGACTGGTGTAAAGACGGGAAAGCAGCTTGGAGCAAAATTTCTAGATAT 667
      |||
1211 GGTAACCGAGGAGTGGTGTGAAGCATTTGATCCTCGCGAGTTTACTTTGAGAAACCGCAT 1270
      |||

```

```

666 GGAACACCGGCTGTAGTTGTAAAGTATTGTATCTCTCGAAGAGTATATGTCAAAAGTCCAGAT 607
      |||
1271 GAAGTAGAATTTAAAGCCCATCAAAAGTAAGCCATCCAGGTCCAAAGCGTGAGGAATATGCC 1330
      |||
606 GAGGTGCAATTTTGAAGTGTAAAGTGTCCAGTAAAGGTCCATTAGAGAGAGATATGCA 547
      |||
1331 TGGTATCAGTCAGCTGTTTAATGGTGGACGAGAGGAGAGCCCATCGAGCTTATGAACATA 1390
      |||
546 TGGTACCAGTCACAGTAGTAAGTGGTGGCGGTGATAGCGCGCTATAGGAGCATATGAACATA 487
      |||
1391 GCTAAGGCTGTTGAGGAACCTGGAGCTGGAGAGAAATATTATTGAACATGCAATGTTGAT 1450
      |||
486 GCAAAAGCTGTCTGAAGAAATTGGCGCGCAGGAGAAATACTACTCAACTGCAATGTTGAT 427
      |||
1451 GGTCAAGGAAAGGATTTCGATATAGATCTGATCAAGCTAATATCCGATGCTGTGAACATT 1510
      |||
426 GGTCAAGGCGGTGGTGTGATATAGATTTGATCAAAATGGTGTGGATGCTGTGACATTT 367
      |||
1511 CCTGTTATCGCAAGCAGCGGTGCAGGAGTCGCTGATCATTCTCCGAAGTCTTTAATGAA 1570
      |||
366 CTGTGCTATTGCAAGCAGTGTGCTGCTGCTCGTGGAGACATTTTCTGAAGTCTTTGAGAA 307
      |||
1571 ACCAAGCATCTGCTGCCCTTCGAGCTGGCATTTTCCATCGCMAAGAGGTCCAAATTAAG 1630
      |||
306 ACAACGCTTCTGCTGCCCTTGTCTGTCGCAATTTTCCATCGGAAAGAGGTTCCTATTCTA 247
      |||
1631 GCTGTTAAAGACGACCTTGTGGAAGGAGGATTGAAGTTAGATTGTA 1677
      |||
246 GCAGTGAAGAGCATCTAGTAGATTTCTGTTGGAGGTGAGAAATGCA 200
      |||

```

RESULT 11
 BM086170
 LOCUS
 DEFINITION
 Gm-c1036-3669 5' similar to TR:Q9SZ30 Q9SZ30 GLUTAMINE
 AMIDOTRANSFERASE/CYCLASE. i, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 BM086170.1 GI:16996807
 EST.
 Glycine max (soybean)
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE
 AUTHORS
 1 (bases 1 to 591)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna
 , A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 , Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 , R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 , R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished

TITLE
 JOURNAL
 COMMENT
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact: cu@resgen.com web site:
 www.resgen.com
 High quality sequence stop: 428.

FEATURES
 source
 1..591
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-c1036-3669"

```

/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-cl036"
/notes="vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies SuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restrictions site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli Electromax DH10B host cells.
This library was constructed in the laboratory of Dr. Lilla
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. E-mail: l-vodkin@uiuc.edu"
91 C 149 q 172 t
179 a

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichoraceae; Lactuca.

I (bases 1 to 716)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://comgenomics.ucdavis.edu/>

Unpublished

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Assumundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
singleton, see <http://cgdpb.ucdavis.edu/> for details.
Plate: OGD7 row: I column: 06

REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1985, Vol. 10, No. 2, pp. 1-15
3. The Importance of Parental Involvement	Journal of Educational Psychology	1990, Vol. 82, No. 3, pp. 1-12
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1995, Vol. 98, No. 4, pp. 1-18
5. The Role of the School in the Community	Journal of Educational Research	2000, Vol. 103, No. 5, pp. 1-20
6. The Impact of Teacher Education on Student Learning	Journal of Educational Research	2005, Vol. 108, No. 6, pp. 1-25
7. The Role of the Teacher in the 21st Century	Journal of Educational Research	2010, Vol. 113, No. 7, pp. 1-30
8. The Impact of Teacher Professional Development on Student Achievement	Journal of Educational Research	2015, Vol. 118, No. 8, pp. 1-35
9. The Role of the Teacher in the Future of Education	Journal of Educational Research	2020, Vol. 123, No. 9, pp. 1-40
10. The Impact of Teacher Leadership on Student Achievement	Journal of Educational Research	2025, Vol. 128, No. 10, pp. 1-45

FEATURES

533 GAATGGTTAGAGTGCCTCATATTGGCTGGAATGCTTTACAAATCAAAAAGATTCACTAGT 592
 475 AATTTTAGATGATCGGGGATCAACATGCTGATTTTGTTCACATCATATCGAGC---CAA 531
 593 TATTTTAGATGATATTCGAAACCCGCAATGATATTTTGTTCATTTCTTATCGTGTGTACC 652
 532 TGCCGAGGACAAACAAGAGTGGATTTCATCTACATGACGCTATGCTGACGATTTTATTGCG 591
 653 GTCTGAGAAAATAAGATTGGCTTTTCATCGACCTGCACTATGGAATTGACTTCATATC 712
 592 ATC 594
 713 ATC 715

 RESULT 13
 CNS06GQF 1041 bp DNA linear GSS 30-NOV-2001
 LOCUS T3 end of clone AS0AA006E12 of library AS0AA from strain CLIB 533
 DEFINITION of Saccharomyces bayanus, genomic survey sequence.
 ACCESSION AL398029.1 GI:12151170
 VERSION
 KEYWORDS
 SOURCE GSS.
 ORGANISM Saccharomyces bayanus
 Saccharomyces bayanus
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 1041)
 Souciet, J. L., Aigle, M., Artiguenave, F., Blandin, G.,
 Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
 Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
 Saurin, P., Tekata, F., Toffano-Nioche, C., Wesolowski-Jouvel, M.,
 Wincker, P. and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEMS Lett. 487 (1), 3-12 (2000)
 20584711
 1152876
 2 (bases 1 to 1041)
 Bon, E., Neuvéglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,
 Aigle, M. and Durrens, P.
 Genomic exploration of the hemiascomycetous yeasts: 5.
 Saccharomyces bayanus var. uvarum
 FEMS Lett. 487 (1), 37-41 (2000)
 20584715
 1152880
 3 (bases 1 to 1041)
 Genoscope.
 Direct Submission
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 Location/Qualifiers
 1. 1041
 /organism="Saccharomyces bayanus"
 /mol_type="genomic DNA"
 /strain="CLIB 533"
 /variety="uvarum"
 /db_xref="taxon:4931"
 /clone="AS0AA006E12"
 /clone_lib="AS0AA"

FEATURES
 source


```

RESULT 14
BE353528
LOCUS
DEFINITION
  BE353528 tomato flower buds 0-3 mm, Cornell University
  EST151725 tomato flower buds 0-3 mm, Cornell University
  Lycopersicon esculentum cDNA clone cTOA1807, mRNA sequence.
ACCESSION
  BE353528
VERSION
  BE353528.1 GI:9291420
KEYWORDS
  EST.
SOURCE
  Lycopersicon esculentum (tomato)
  Lycopersicon esculentum
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
  1 (bases 1 to 569)
REFERENCE
  van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
  ,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,
  Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  Generation of ESTs from tomato flower tissue, 0-3 mm buds
  Unpublished
  Contact: CUGI
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
FEATURES
  source
    1..569
    /location/Qualifiers
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="TA496"
    /db_xref="taxon:4081"
    /clone="cTOA1807"
    /tissue_type="flower"
    /dev_stage="0-3mm buds"
    /clone_lib="tomato flower buds 0-3 mm, Cornell University"
    /notes="vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; supplier: Tanksley; Flower buds and flowers were
    taken from greenhouse plants (4-8 wks old, TA496). They
    were immediately frozen in liquid nitrogen and then
    size-separated while remaining frozen."
  BASE COUNT
    156 a 113 c 136 g 164 t
    ORIGIN
      Query Match 15.3%; Score 298.4; DB 10; Length 569;
      Best Local Similarity 73.3%; Pred. No. 2.9e-27;
      Matches 396; Conservative 0; Mismatches 141; Indels 3; Gaps 1;

QY 55 CGCTTGGCTGGTGCAGCGGAGATTCTGTGGTGACTTACTTGATTACGGTCTGGAAA 114
    |||||
DB 30 CGCTTCCGCCACCAAGTCCCGATGATCTGTGGTAACTCTACTTGATTGCTGGCAA 89

QY 115 TGTAGGAGTGTGAGGAACCCATCCGCACATTTGGATTGATATCAAGATGTCGAAA 174
    |||||
DB 90 TGTGCGGAGTGTGAGGAATCAATTAATATCTCGGTTTCGACATCAAGATGTCAAA 149

QY 175 GCCAGAGGATATTTCTAAATGCTAAGCGCCCTTATCTTCTCGGGGTTGGGCGCTTGCACC 234
    |||||
DB 150 ACCACAGGACATTTTGAAGCAAAACGACTCATTTTCTCGGGGTTGGTGTCTTGTCTCC 209

QY 235 TGCATAGGATGTTCTTATTTCGTAAGAGGCTGGCTGAAGCACTCTGTACTTACATTCAGAA 294
    |||||
DB 210 TGCCATGGATGTTCTTAAACAAGAAGGAATGGCTGAAGCACTTTGTGCTTACATTCAGCA 269

QY 295 TGATCGACCTTCTCTGGGTATATCGCTGGGATTCGAGTACTCTTTGAGTCAAGTGAAGA 354
    |||||
DB 270 AGATCGCCCATCTCTAGGCAATTTGTCTTGACTGCAGTACTCTTTTGGTCAAGTGAAGA 329

QY 355 AAATGTTCCAAATTCAGGCTCTCGCTTGAATTCCTGGACGGGTTGGGCGTTTTCGAATCATC 414
    |||||
DB 330 AAATGGAACAAGTAAGGGTCTTGTTGATTCCTGGAGTAGTTGGACGTTTTCGACTCTTC 399

QY 415 CAATGGTTTAAAGGGTGCACATATTTGGATGGCATGCCCTTGGATATAAAGGAAGGGTCAGC 474
    |||||

```

```

Db 390 TAAAGGATCAGAGTACCTCATTGGCTGGAAATGCATTGGAATAGCAAAAGATACTCA 449
QY 475 AATTTTAGATGATGTGGGAAATCAACATGTGTATTGTTTCTACTCATATCGAGCCA---A 531
    |||||
Db 450 AATTTTGGATGATGTGGGAAATTCACATGTCTTATTTCGTACATTTCTTATCGGCATGCC 509
    |||||
QY 532 TGCCGAGGACAAACAAGAGTGGATTTCATCTACATGCAGCTATGTGTGAGATTTTATGTC 591
    |||||
Db 510 GTCAGATGAATAAGAGAGTGGATATCATCGACTGCAAGATATGTTGGTGGATTCATAGC 569

RESULT 15
BG833390/c
LOCUS
DEFINITION
  BG833390 x2 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA,
  mRNA sequence.
ACCESSION
  BG833390
VERSION
  BG833390.1 GI:14191160
KEYWORDS
  EST.
SOURCE
  Zea mays
  Zea mays
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
  clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 511)
REFERENCE
  Maize ESTs from various cDNA libraries sequenced at Stanford
  University
  Unpublished
  Contact: Walbot V
  Department of Biological Sciences
  Stanford University
  855 California Ave, Palo Alto, CA 94304, USA
  Tel: 650 723 2227
  Fax: 650 725 8221
  Email: walbot@stanford.edu
  Plate: 951007 row: E column: 08.
FEATURES
  source
    1..511
    /organism="Zea mays"
    /mol_type="mRNA"
    /cultivar="BMS (Black Mexican Sweet)"
    /db_xref="taxon:4577"
    /tissue_type="suspension culture"
    /dev_stage="mixed logarithmic and stationary growth
    phases"
    /lab_host="DH10B"
    /clone_lib="951 - BMS tissue from Walbot Lab (GR)"
    /note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
    library was prepared by George Rudenko using poly (A)
    selected RNA and Universal Riboclone cDNA Synthesis System
    (Promega). cDNA was synthesized using both random and
    oligo(dT) primers in separate reactions and equipped with
    EcoRI adaptors. Library was size-fractionated on agarose
    gels (for insert size >400bp) and non-directionally cloned
    into EcoRI-digested pUC19 vector. Blue/white selection on
    carbenicillin-containing plates was used to recover
    positive clones."
  BASE COUNT
    138 a 144 c 84 g 145 t
    ORIGIN
      Query Match 14.8%; Score 289; DB 12; Length 511;
      Best Local Similarity 73.3%; Pred. No. 4.3e-26;
      Matches 370; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 1174 AGGAAAGACAGCATAGACAGATATCTACAGTATATGTAACAGGAGTGGTGTAAAG 1233
    |||||
Db 511 AGGAAAGACAGCTTGGACAAATCTCTAGAGTATATGCAATCAGGCTGTAGTGTCTCAG 452

QY 1234 CATTTGATCTCTCCGAGTTTACTTGAGAAAACCCGATGAAGTAGAATTTAAAGCCATCAA 1293
    |||||
Db 451 TATTGATCTCTCAGCGGTATATGTCAAAGATGTGCCATTTTAAACTGTAAA 392

```

```
Qy 1294 AGTAAGCATCCAGGTCCAAACGGTGAGGAATATGCTGGTATCAGTGCACACTGTTAATGG 1353
Db 391 GGTGTCACCTAAAGTCCATCGGAGAGAAATATGCATGGTACCACTGCACACTGATGG 332
Qy 1354 TGGACGAGAAAGGAGAGCCCATCGGAGCTTATGAACCTAGCTAAGGCTGTTGAGGAACTTGG 1413
Db 331 TGGACGTGATAGCCGAGCTATAGGAGCATATGAACCTAGCGAAAGCTGTGGAGAAATTGGG 272
Qy 1414 AGCTGGAGAAATATATTGAACCTGCATTGATGCTCAAGGAAAGGATTCGATAT 1473
Db 271 CGCAGGAGAAATATCTTAACTGCATTGATGCTGATGGCCAAAGGTTGTGGATTTGACAT 212
Qy 1474 AGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCTGTTATCGAAGCAGCGGTGC 1533
Db 211 AGATTTGGTTAAATGGTTTCTGATGCTGTGACAACTCCCTGTCTTGGAGCAGTGGTGC 152
Qy 1534 AGGAGTCGCTGATCAGCTTCTCCGAAGTCTTTAATGAACCAACGCACTGCTGCCCTTGC 1593
Db 151 TGGAGCTGTTCAACATTTTCTGAATTTTGGAGAAACAAATGCTTCTGCTGCCCTTGC 92
Qy 1594 AGCTGGCATTTTCCATCGAAAGGTTCCAAATTAAGGCTGTTAAGAGCACTTGTGAA 1653
Db 91 TGCTGGCATTTTCCACCGGAAAGAGGTTCCCTATATACTAGCAGTGAAGAGCATCTGGTCAA 32
Qy 1654 GGAAGGGATTGAAGTTAGATTGTAA 1678
Db 31 TGCTGGTGGAGGTCAGGGGTGTA 7
```

Search completed: January 21, 2004, 02:45:24
Job time : 4253 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:20:02 ; Search time 46 Seconds
(without alignments)
1925.422 Million cell updates/sec

Title: US-09-831-233A-2
Perfect score: 2877
Sequence: 1 HEKELASTKPFQNGFRIRAL.....EVPKAVKHEHLKEGIEVRL 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A_Geneseq_19Jun03.*			
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2877	100.0	558	21	Impatiens balsamia
2	2284.5	79.4	592	23	Herbicidally activ
3	2160	75.1	593	21	Arabidopsis thalia
4	2157	75.0	593	21	Cress mutant gluta
5	2157	75.0	593	21	Cress mutant cycla
6	2154	74.9	593	21	Cress mutant gluta
7	2154	74.9	593	21	Cress mutant cycla
8	2154	74.9	593	21	Cress mutant cycla
9	2154	74.9	593	21	Cress mutant cycla

10	2154	74.9	593	21	AA94244	Cress mutant cycla
11	2154	74.9	593	21	AA94246	Cress mutant cycla
12	2152	74.8	593	21	AA94239	Cress mutant cycla
13	2152	74.8	593	21	AA94240	Cress mutant cycla
14	2152	74.8	593	21	AA94243	Cress mutant cycla
15	2152	74.8	593	21	AA94245	Cress mutant cycla
16	2151	74.8	593	21	AA94232	Cress mutant gluta
17	2150	74.7	593	21	AA94233	Cress mutant gluta
18	2150	74.7	593	21	AA94234	Cress mutant gluta
19	1279	44.5	295	21	AA94226	Corn putative glut
20	630	21.9	141	21	AA94230	Soybean putative g
21	481.5	16.7	258	22	AA92037	C glutamicum prote
22	481.5	16.7	258	22	AA92037	C glutamicum prote
23	467.5	16.2	251	23	ABB48751	Listeria monocytog
24	439	15.3	256	23	ABP65662	Bifidobacterium lo
25	430.5	15.0	258	23	ABP39802	staphylococcus epi
26	428	14.9	255	24	ABP77948	N. gonorrhoeae ami
27	415.5	14.4	256	22	AAU51170	Propionibacterium
28	413.5	14.4	247	22	AA981599	S. epidermidis ope
29	409	14.2	208	23	ABB48753	Listeria monocytog
30	399.5	13.9	259	23	ABB54545	Lactococcus lactis
31	357	12.4	192	20	AAW97360	An amidotransferas
32	357	12.4	192	20	AAW97361	Amidotransferase d
33	355	12.3	192	23	ABP39807	Staphylococcus epi
34	339.5	11.8	222	22	AAU63592	Propionibacterium
35	339	11.8	226	24	ABP77949	N. gonorrhoeae ami
36	338	11.7	202	23	ABB54543	Lactococcus lactis
37	336.5	11.7	215	23	ABP66036	Bifidobacterium lo
38	332	11.5	211	22	AA92040	C glutamicum prote
39	332	11.5	211	22	AA92040	C glutamicum prote
40	286	9.9	114	23	ABP3635	Corynebacterium gi
41	285	9.9	130	21	AA94225	Human ORF2608 prot
42	284	9.9	158	22	AA94225	Corn putative glut
43	264	9.2	251	18	AA937598	S. epidermidis ope
44	262	8.4	58	21	AAW37558	Hsp protein invol
45	215	7.5	127	22	AA94229	Soybean putative g
						S. epidermidis ope

ALIGNMENTS

RESULT 1	
AA94224	AA94224 standard; Protein; 558 AA.
ID	AA94224 standard; Protein; 558 AA.
XX	
AC	AA94224;
XX	
DT	08-AUG-2000 (first entry)
XX	
DE	Impatiens balsamia glutamine amidotransferase.
XX	
KW	Glutamine amidotransferase; histidine biosynthesis; herbicide;
KW	fungicide.
XX	
OS	Impatiens balsamia.
XX	
PN	WO200028053-A2.
XX	
PD	18-MAY-2000.
XX	
PF	04-NOV-1999; 99WO-US25950.
XX	
PR	05-NOV-1998; 98US-0107275.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E I.
XX	
PI	Allen SM, Huang LL, Falco SC, Rafalski AJ;
XX	
DR	WPI: 2000-376564/32.
DR	N-PSDB; AA27325.
XX	
PT	Novel polynucleotides encoding plant glutamine amidotransferase
PT	homologues, useful for producing transgenic plants and as probes or

```

PT primers -
XX
PS Claim 10; Page 41; 53pp; English.
XX
CC The present sequence is the Impatiens balsamia glutamine
CC amidotransferase protein sequence. The protein forms a crucial stage
CC in the histidine biosynthesis pathway. The gene sequence can be
CC used to create transgenic plants which express different amounts of the
CC protein, to identify loss of function mutants and to produce the
CC protein in a host cell, for example a bacterium. The protein can be used
CC to identify inhibitors which may be useful as fungicides and herbicides.
XX
SQ Sequence 558 AA;

Query Match 100.0%; Score 2877; DB 21; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.5e-252;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEKELASTKPGNGFRIRAAAGAGGDSVVTLLDYGAGNVRSVRNAIRTLGFDIKDVQKPE 60
DB 1 HEKELASTKPGNGFRIRAAAGAGGDSVVTLLDYGAGNVRSVRNAIRTLGFDIKDVQKPE 60

QY 61 DIINAKELIIPFGVGAAPAMDVILIRKGLAEALCTYIQNDPFLGICLGLQLLPESSENG 120
DB 61 DIINAKELIIPFGVGAAPAMDVILIRKGLAEALCTYIQNDPFLGICLGLQLLPESSENG 120

QY 121 PIQGLGLIPGRVGRFESSNGLRVPHIGWHDLDIKESAILDDVGNQHVYFVHSYRANAED 180
DB 121 PIQGLGLIPGRVGRFESSNGLRVPHIGWHDLDIKESAILDDVGNQHVYFVHSYRANAED 180

QY 181 NKEWISSTCSYGDDFTASIQGNVHAVQFHPKSGGVGLSILRRFLNADSFNNKROKPMN 240
DB 181 NKEWISSTCSYGDDFTASIQGNVHAVQFHPKSGGVGLSILRRFLNADSFNNKROKPMN 240

QY 241 GKASKLAKRVIACLDVANDNGDLVVTGQDQYDVRTEENEVRNLGKPVLAGQYVLDG 300
DB 241 GKASKLAKRVIACLDVANDNGDLVVTGQDQYDVRTEENEVRNLGKPVLAGQYVLDG 300

QY 301 ADEVSFLNIITGPRDFPLGDLPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVA 360
DB 301 ADEVSFLNIITGPRDFPLGDLPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVA 360

QY 361 SEVFRSGADKVSIGSDAVYTAEEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLR 420
DB 361 SEVFRSGADKVSIGSDAVYTAEEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLR 420

QY 421 KPDEVEPKAIKVSHPGNGBEYAWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCI 480
DB 421 KPDEVEPKAIKVSHPGNGBEYAWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCI 480

QY 481 DCDGQGGKGFIDILKILSDAVNIPVIASSGAGVADHSEVFNETNASAAALAGIFHRKEV 540
DB 481 DCDGQGGKGFIDILKILSDAVNIPVIASSGAGVADHSEVFNETNASAAALAGIFHRKEV 540

QY 541 PIKAVKEHLKKEGIEVRL 558
DB 541 PIKAVKEHLKKEGIEVRL 558

RESULT 2
ABB93115
ID ABB93115 standard; Protein; 592 AA.
XX
AC ABB93115;
XX
31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2326.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX

```

```

PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-BF09892.
XX
PR 28-AUG-2001; 2001WO-BF09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-369010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
XX Claim 5; SEQ ID NO 2326; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 592 AA;

Query Match 79.4%; Score 2284.5; DB 23; Length 592;
Best Local Similarity 80.7%; Pred. No. 1.2e-198;
Matches 435; Conservative 56; Mismatches 45; Indels 3; Gaps 2;

QY 21 AGAGGDSVVTLLDYGAGNVRSVRNAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 AASTSDSVVTLLDYGAGNVRSIRNARHLGFSIKDVQTFDILNADRLIFPGVGAFAPAM 115

QY 81 DVLIRKGLAEALCTYIQNDPFLGICLGLQLLPESSENGPIQGLGIFGRVGRFESSNG 140
DB 116 DVLNRTGMAEALCKYIENDRPFLGICLGLQLLPDSSENGPVKGLGVIPIGVGRFDASAG 175

QY 141 LRPHIGWHDLDIKESAILDDVGNQHVYFVHSYRANAEDNKEWISSTCSYGDDFTASI 199
DB 176 IRVPHIGWHDLDIKESAILDDVGNQHVYFVHSYRAIPSDENKOWISSTCSYGESFISI 235

QY 200 QKGNVHAVQFHPKSGGVGLSILRRFLNADSFNNKROKPMNGKASKLAKRVIACLDVAN 259
DB 236 RRGNVHAVQFHPKSGGVGLSILRRFLNADSFNNKROKPMNGKASKLAKRVIACLDVAN 293

QY 260 DNGDLVVTGQDQYDVRTEENEVRNLGKPVLAGQYVLDGADDEVFLNITGPRDFPLGD 319
DB 294 DKGDLVVTGQDQYDVRTEENEVRNLGKPVLAGQYVLDGADDEVFLNITGPRDFPLGD 353

QY 320 LPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVASSEYFSGADKVSIGSDAVY 379
DB 354 LPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVASSEYFSGADKVSIGSDAVY 413

QY 380 TAEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEVEFKAIKVSHPGNG 439
DB 414 AAEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEVEFKAIKVSHPGNG 473

QY 440 EEWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCIDCDGQGGKGFIDILKILSD 499
DB 474 EEWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCIDCDGQGGKGFIDILKILSD 533

QY 500 ANNIPVIASSGAGVADHSEVFNETNASAAALAGIFHRKEVPIKAVKEHLKKEGIEVRL 558
DB 534 SVGIPVIASSGAGTDPHFSEVFNETNASAAALAGIFHRKEVPIKAVKEHLKKEGIEVRL 592

```

```
RESULT 3
AA94231
ID AAY94231 standard; Protein; 593 AA.
XX
AC AAY94231;
XX
DT 08-AUG-2000 (first entry)
XX
DE Arabidopsis thaliana HisH enzyme.
XX
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis.
XX
OS Arabidopsis thaliana.
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Huang LL, Falco SC, Rafalski AJ;
XX
DR WPI; 2000-376564/32.
XX
PT Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
PT primers
XX
PS Example 9; Page 50-52; 53pp; English.
XX
CC The present sequence is the mouse ear cress HisH enzyme, which is
CC involved in histidine biosynthesis. It was used in site-directed
CC mutagenesis to find the active sites within the enzyme, which is made up
CC of a cyclase and a glutamine amidotransferase. This is useful for the
CC identification of herbicides and fungicides which inhibit the enzyme.
XX
SQ Sequence 593 AA;
Query Match 75.1%; Score 2160; DB 21; Length 593;
Best Local Similarity 76.7%; Pred. No. 2.5e-187;
Matches 414; Conservative 60; Mismatches 62; Indels 4; Gaps 3;
QY 21 AGAGDSVVTLLDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 ASSTSDSVVTLDDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 115
QY 81 DVLIRKGLAALCTYQNDPFLGICLGLQLPESSEENGPIQGLGILPGVRORPSSNG 140
DB 116 DVLNRTGMALCKYIENDRPFLGICLGLQLPESSEENGPIQGLGILPGVRORPSSNG 175
QY 141 LRVPHIGWALDIKEGSAIILDDVGNHVVYFVHVSRA-NAEDNKEWISSTCSYGDDETIAS 199
DB 176 IRVPHIGWALDIKEGSAIILDDVGNHVVYFVHVSRAI-PSDENKDWISSTCSYGFISFI 235
QY 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKRAKVIACLDVRAN 259
DB 236 RRGNNVAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKRAKVIACLDVRAN 293
QY 260 DNGDLVVTGQDVDRERTENEVRNLGKVELAGQYVLDGAEVDFNLITGPRDPLGD 319
DB 294 DKGDLVVTGQDVDRERTENEVRNLGKVELAGQYVLDGAEVDFNLITGPRDPLGD 353
QY 320 LPMIQLVLRASENVFVPLTGGGIRDTDANGRYSSLEVASFVRSKADKVISGDVAV 379
DB 354 LPMIQLVLRASENVFVPLTGGGIRDTDANGRYSSLEVASFVRSKADKVISGDVAV 413
QY 380 TABEYIKTKGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEVEFKAIKVSHFGPNG 439
DB 414 AAEFFIKSGVTKGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEVEFKAIKVSHFGPNG 473
QY 440 EYAWYQCTVNGRGRRPIGAYELAKAVEELGAGELLNCIDCGGKGKGFDDILKLI 499
DB 474 EYAWYQCTVNGRGRRPIGAYELAKAVEELGAGELLNCIDCGGKGKGFDDILKLI 533
QY 500 AVNIPVIASSGAGVADHFSEVFNETNASAAAGIF-HRKEVPIKAVKSHLLKEGIEVRL 558
DB 534 SVGIPVIASSGAGTDPHFSEVFEEDKRICRACRHPFPPRGYQSQSVKSHLQEEIEVRI 593
RESULT 4
AA94236
ID AAY94236 standard; Protein; 593 AA.
XX
AC AAY94236;
XX
DT 08-AUG-2000 (first entry)
XX
DE Cress mutant glutamine amidotransferase, E248Q substitution.
XX
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis; mutein; substitution;
KW mutant.
XX
OS Arabidopsis thaliana.
XX
PN Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 248 /note= "wild-type Glu substituted by Gln"
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Huang LL, Falco SC, Rafalski AJ;
XX
DR WPI; 2000-376564/32.
XX
PT Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
PT primers
XX
PS Example 9; Page -; 53pp; English.
XX
CC The present sequence is a mutant of the mouse ear cress HisH enzyme,
CC which is involved in histidine biosynthesis. It was produced by
CC site-directed mutagenesis, which was carried out to find the active
CC sites within the enzyme, which is made up of a cyclase and a glutamine
CC amidotransferase. The mutation in this case is in the glutamine
CC amidotransferase domain. Determining the active sites is useful for the
CC identification of herbicides and fungicides which inhibit the enzyme.
CC This residue was shown to form part of the active site.
CC Note: The present sequence is not shown in the specification but is
CC derived from the Arabidopsis thaliana wild-type His HF sequence given in
CC the sequence listing (see AAY94231).
XX
SQ Sequence 593 AA;
Query Match 75.0%; Score 2157; DB 21; Length 593;
Best Local Similarity 76.5%; Pred. No. 4.7e-187;
Matches 413; Conservative 61; Mismatches 62; Indels 4; Gaps 3;
QY 21 AGAGDSVVTLLDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 ASSTSDSVVTLDDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 115
QY 81 DVLIRKGLAALCTYQNDPFLGICLGLQLPESSEENGPIQGLGILPGVRORPSSNG 140
DB 116 DVLNRTGMALCKYIENDRPFLGICLGLQLPESSEENGPIQGLGILPGVRORPSSNG 175
QY 141 LRVPHIGWALDIKEGSAIILDDVGNHVVYFVHVSRA-NAEDNKEWISSTCSYGDDETIAS 199
DB 176 IRVPHIGWALDIKEGSAIILDDVGNHVVYFVHVSRAI-PSDENKDWISSTCSYGFISFI 235
QY 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKRAKVIACLDVRAN 259
DB 236 RRGNNVAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKRAKVIACLDVRAN 293
QY 260 DNGDLVVTGQDVDRERTENEVRNLGKVELAGQYVLDGAEVDFNLITGPRDPLGD 319
DB 294 DKGDLVVTGQDVDRERTENEVRNLGKVELAGQYVLDGAEVDFNLITGPRDPLGD 353
QY 320 LPMIQLVLRASENVFVPLTGGGIRDTDANGRYSSLEVASFVRSKADKVISGDVAV 379
DB 354 LPMIQLVLRASENVFVPLTGGGIRDTDANGRYSSLEVASFVRSKADKVISGDVAV 413
QY 380 TABEYIKTKGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEVEFKAIKVSHFGPNG 439
DB 414 AAEFFIKSGVTKGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEVEFKAIKVSHFGPNG 473
QY 440 EYAWYQCTVNGRGRRPIGAYELAKAVEELGAGELLNCIDCGGKGKGFDDILKLI 499
DB 474 EYAWYQCTVNGRGRRPIGAYELAKAVEELGAGELLNCIDCGGKGKGFDDILKLI 533
QY 500 AVNIPVIASSGAGVADHFSEVFNETNASAAAGIF-HRKEVPIKAVKSHLLKEGIEVRL 558
DB 534 SVGIPVIASSGAGTDPHFSEVFEEDKRICRACRHPFPPRGYQSQSVKSHLQEEIEVRI 593
```

Db 56 ASSTDSVVTLLDYGAGNVRISRNALRHGFSIKDVQTPGDIILNADRLIPPGVGPAPAM 115
 Qy 81 DVLIRKGLAALCTYIONDRPFLGICLGLQLLPESSENGPIQGLGLIPGRVGRFPSSNG 140
 Db 116 DVLNRTGMALCKYIENDRPFLGICLGLQLLPDSSENGPVKGLGVPVIGVGRFPDASAG 175
 Qy 141 LRVPHICGHALDIKEGSAIILDDVGNQHVYFVHSYRA-NAEDNKESWISSTCSYGGDDFTASI 199
 Db 176 IRVPHIGWALQVKGSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYSGESPTSI 235
 Qy 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKRPQKPMNGKASKLAKRVIACLDVRAN 259
 Db 236 RRGNVHAVQHPKSGVGLSVLRFLHPKL--PATQKPMEGKASKLAKRVIACLDVRTN 293
 Qy 260 DNGDLVVTGQDQYDVRTEENVRNLGKPVELAGQYILDGADSEVFLNITGFRDFPLGD 319
 Db 294 DKGDLVVTGQDQYDVRTEENVRNLGKPVELAGQYILDGADSEVFLNITGFRDFPLGD 353
 Qy 320 LPMLOVLQASENVFVPLTVGGGIRDTDANGRYYSLEVAEYFRSGADKVSIGSDAVY 379
 Db 354 LPMLOVLQRTSKNVFVPLTVGGGIRDTDANGRYYSLEVAEYFRSGADKVSIGSDAVF 413
 Qy 380 TABEYIKTVTKGKSSIEQISTVYGNQAVVVSIDPRVYLKPKDEVEFKAIVKSHPGPNG 439
 Db 414 AAEFIKSGVTKGKSSLEQISRYVYGNQAVVVSIDPRVYVNHDPDVPYKIVRTNPGNG 473
 Qy 440 EEWYQCTVNGRGREGPIGAYELAKAVEELGAGEILLNCIDCGQKGFIDILKILSD 499
 Db 474 EEWYQCTVSGQGEGRPIGAFELAKAVEELGAGEILLNCIDCGQKGFIDILKILSD 533
 Qy 500 AVNIPVTASSGAGVADHPSEVFNETNASALAAGIF-HRKEVPKAVKEHLKEGIEVRL 558
 Db 534 SVGIPVIASSGAGTPDHPSEVFEEDEKRICRACCRHPPPERGYQSQSVKEHLQERIEVRI 593

RESULT 5

AA94242
 ID AA94242 standard; Protein; 593 AA.
 AC AA94242;
 XX 08-AUG-2000 (first entry)
 XX Cress mutant cyclase, S409A substitution.
 XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
 KW mouse ear cress; site-directed mutagenesis; muten; substitution;
 KW mutant.
 XX Arabidopsis thaliana.
 OS Synthetic.
 FT Key Location/Qualifiers
 FT Misc-difference 409 /note= "Wild-type Ser substituted by Ala"
 FT
 FN W0200028053-A2.
 XX
 XX 18-MAY-2000.
 XX
 XX 04-NOV-1999; 99WO-US25950.
 XX
 XX 05-NOV-1998; 98US-0107275.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Allen SM, Huang LL, Falco SC, Rafalski AJ;
 XX
 XX WPI; 2000-376564/32.
 XX
 XX Novel polynucleotides encoding plant glutamine amidotransferase
 PT homologues, useful for producing transgenic plants and as probes or
 PT primers -

XX Example 9; Page -; 53pp; English.
 PS The present sequence is a mutant of the mouse ear cress HisHFP enzyme,
 CC which is involved in histidine biosynthesis. It was produced by
 CC site-directed mutagenesis, which was carried out to find the active
 CC sites within the enzyme, which is made up of a cyclase and a glutamine
 CC amidotransferase. The mutation in this case is in the cyclase
 CC domain. Determining the active sites is useful for the identification of
 CC herbicides and fungicides which inhibit the enzyme. This residue is
 CC not expected to be within the active site.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from the Arabidopsis thaliana wild-type His HF sequence given in
 CC the sequence listing (see AA94231).

XX Sequence 593 AA;

Query Match 75.0%; Score 2157; DB 21; Length 593;
 Best Local Similarity 76.5%; Pred. No. 4.7e-187;
 Matches 413; Conservative 61; Mismatches 62; Indels 4; Gaps 3;

Qy 21 AGAGDSVVTLLDYGAGNVRISRNALRHGFSIKDVQTPGDIILNADRLIPPGVGPAPAM 80
 Db 56 ASSTDSVVTLLDYGAGNVRISRNALRHGFSIKDVQTPGDIILNADRLIPPGVGPAPAM 115
 Qy 81 DVLIRKGLAALCTYIONDRPFLGICLGLQLLPESSENGPIQGLGLIPGRVGRFPSSNG 140
 Db 116 DVLNRTGMALCKYIENDRPFLGICLGLQLLPDSSENGPVKGLGVPVIGVGRFPDASAG 175
 Qy 141 LRVPHICGHALDIKEGSAIILDDVGNQHVYFVHSYRA-NAEDNKESWISSTCSYGGDDFTASI 199
 Db 176 IRVPHIGWALQVKGSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYSGESPTSI 235
 Qy 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKRPQKPMNGKASKLAKRVIACLDVRAN 259
 Db 236 RRGNVHAVQHPKSGVGLSVLRFLHPKL--PATQKPMEGKASKLAKRVIACLDVRTN 293
 Qy 260 DNGDLVVTGQDQYDVRTEENVRNLGKPVELAGQYILDGADSEVFLNITGFRDFPLGD 319
 Db 294 DKGDLVVTGQDQYDVRTEENVRNLGKPVELAGQYILDGADSEVFLNITGFRDFPLGD 353
 Qy 320 LPMLOVLQASENVFVPLTVGGGIRDTDANGRYYSLEVAEYFRSGADKVSIGSDAVY 379
 Db 354 LPMLOVLQRTSKNVFVPLTVGGGIRDTDANGRYYSLEVAEYFRSGADKVSIGSDAVF 413
 Qy 380 TABEYIKTVTKGKSSIEQISTVYGNQAVVVSIDPRVYLKPKDEVEFKAIVKSHPGPNG 439
 Db 414 AAEFIKSGVTKGKSSLEQISRYVYGNQAVVVSIDPRVYVNHDPDVPYKIVRTNPGNG 473
 Qy 440 EEWYQCTVNGRGREGPIGAYELAKAVEELGAGEILLNCIDCGQKGFIDILKILSD 499
 Db 474 EEWYQCTVSGQGEGRPIGAFELAKAVEELGAGEILLNCIDCGQKGFIDILKILSD 533
 Qy 500 AVNIPVIASSGAGVADHPSEVFNETNASALAAGIF-HRKEVPKAVKEHLKEGIEVRL 558
 Db 534 SVGIPVIASSGAGTPDHPSEVFEEDEKRICRACCRHPPPERGYQSQSVKEHLQERIEVRI 593

RESULT 6

AA94235
 ID AA94235 standard; Protein; 593 AA.
 XX
 AC AA94235;
 XX
 XX 08-AUG-2000 (first entry)
 XX Cress mutant glutamine amidotransferase, E248A substitution.
 XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
 KW mouse ear cress; site-directed mutagenesis; muten; substitution;
 KW mutant.
 XX Arabidopsis thaliana.
 OS

Db 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLLFDSSEQNGPVKGLGVPIGVGRFDASAG 175
 Qy 141 LRVPHIGWALDIKEGSAILLDDVGNQVYFVHSYRA--NAEDNKWISSTCSYGGDDFTASI 199
 Db 176 IRVPHIGWALQVKGKSEILLDDVGNRHVYFVHSYRAIPSDENKDWISSTCNGESPTSSI 235
 Qy 200 QKGNHVAQFHPKESGGVLSILRRFLNADSFNKKRQKPMNGKASKLAKRVIACLDVRAN 259
 Db 236 RRGNVHVAQFHPKESGEVGLSVLRFLHPKL--PATQKPMGKASKLAKRVIACLDVRTN 293
 Qy 260 DNGDLVVTGQDYDVRTEENEVRNLGKPVLAGQYLLDGADEVSLNITGFRDPLD 319
 Db 294 DKGDLVVTAGQDYDVRQSENEVRNLGKPVLAGQYLLDGADEISFLNITGFRDPLD 353
 Qy 320 LPMQLVLRASENVFVPLTVGGGIRDTDANGRYYSLEVASSEYFRSGADKVGSDAVY 379
 Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDTDASGRYYSLEVAABEYFRSGADKMGSDAVF 413
 Qy 380 TAEEYIKTGKTKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDVEFEKAIKVSHPGNG 439
 Db 414 AAEFIKSGVKTGKSSLEQISRVYGNQAVVVSIDPRRVYVNHDPDDVPYKVRVTNPGNG 473
 Qy 440 EEWYQCTVNGGREGPIGAYELAKAVEBELGAGEILLNCIDCDGQKGFIDILIKLISD 499
 Db 474 EEWYQCTVSGGQEGRPAGAFELAKAVEBELGAGEILLNCIDCDGQKGFIDILVKLISD 533
 Qy 500 AVNIPVIASSGAGVADHFSEVFNETNASALAAGIP-HRKEVPIKAVEKHELLKEGIEVRL 558
 Db 534 SVGPVVIASSGAGTPDHFSEVFEEDKRICRACCRHPPPERGYQSQVKEHLOEBERIEVRI 593

RESULT 8

AA194238
 ID AAY94238 standard; Protein; 593 AA.

XX AC AAY94238;

DT 08-AUG-2000 (first entry)

XX Cress mutant cyclase, D334A substitution.

XX Glutamine amidotransferase; histidine biosynthesis; cyclase;

XX mouse ear cress; site-directed mutagenesis; mutein; substitution;
 mutant.

XX Arabidopsis thaliana.

OS Synthetic.

PH Key Location/Qualifiers

FT Misc-difference 334 /note= "Wild-type Asp is substituted by Ala"

XX WO200028053-A2.

XX 18-MAY-2000.

XX 04-NOV-1999; 99WO-US25950.

XX 05-NOV-1998; 98US-0107275.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Allen SM, Huang LL, Falco SC, Rafaleki AJ;

XX WPI; 2000-376564/32.

XX Novel polynucleotides encoding plant glutamine amidotransferase
 homologues, useful for producing transgenic plants and as probes or
 primers -

XX Example 9; Page -; 53pp; English.

XX The present sequence is a mutant of the mouse ear cress HisHf enzyme,

CC which is involved in histidine biosynthesis. It was produced by
 site-directed mutagenesis, which was carried out to find the active
 sites within the enzyme, which is made up of a cyclase and a glutamine
 amidotransferase. The mutation in this case is in the cyclase
 domain. Determining the active sites is useful for the identification of
 herbicides and fungicides which inhibit the enzyme. This residue is
 not expected to be within the active site.
 CC Note: The present sequence is not shown in the specification but is
 derived from the Arabidopsis thaliana wild-type HisHf sequence given in
 the sequence listing (see AAY94231). D334A. The residue in the
 CC Note: The mutant is stated as being Gly.
 CC wild-type is stated as being Gly.

XX Sequence 593 AA;

Query Match 74.9%; Score 2154; DB 21; Length 593;

Best Local Similarity 76.5%; Pred. No. 8.8e-187;

Matches 413; Conservative 60; Mismatches 63; Indels 4; Gaps 3;

Qy 21 AGAGGDSVVTLLDYGAGNVRSVRNARTLGFDIKDVKPEDILNAKRLIFPGVGAFAPAM 80

Db 56 ASTSDSVVTLLDYGAGNVRSIRNARHLGFSIKDVQTFDILNADRLIFPGVGAFAPAM 115

Qy 81 DVLIRKGLAEALCTYIQNDPRPFLGICLGLQLLFESSENGPIQGLGIPRVRGPFSSNG 140

Db 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLLFDSSEQNGPVKGLGVPIGVGRFDASAG 175

Qy 141 LRVPHIGWALDIKEGSAILLDDVGNQVYFVHSYRA--NAEDNKWISSTCSYGGDDFTASI 199

Db 176 IRVPHIGWALQVKGKSEILLDDVGNRHVYFVHSYRAIPSDENKDWISSTCNGESPTSSI 235

Qy 200 QKGNHVAQFHPKESGGVLSILRRFLNADSFNKKRQKPMNGKASKLAKRVIACLDVRAN 259

Db 236 RRGNVHVAQFHPKESGEVGLSVLRFLHPKL--PATQKPMGKASKLAKRVIACLDVRTN 293

Qy 260 DNGDLVVTGQDYDVRTEENEVRNLGKPVLAGQYLLDGADEVSLNITGFRDPLD 319

Db 294 DKGDLVVTAGQDYDVRQSENEVRNLGKPVLAGQYLLDGADEISFLNITGFRDPLD 353

Qy 320 LPMQLVLRASENVFVPLTVGGGIRDTDANGRYYSLEVASSEYFRSGADKVGSDAVY 379

Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDTDASGRYYSLEVAABEYFRSGADKMGSDAVF 413

Qy 380 TAEEYIKTGKTKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDVEFEKAIKVSHPGNG 439

Db 414 AAEFIKSGVKTGKSSLEQISRVYGNQAVVVSIDPRRVYVNHDPDDVPYKVRVTNPGNG 473

Qy 440 EEWYQCTVNGGREGPIGAYELAKAVEBELGAGEILLNCIDCDGQKGFIDILIKLISD 499

Db 474 EEWYQCTVSGGQEGRPAGAFELAKAVEBELGAGEILLNCIDCDGQKGFIDILVKLISD 533

Qy 500 AVNIPVIASSGAGVADHFSEVFNETNASALAAGIP-HRKEVPIKAVEKHELLKEGIEVRL 558

Db 534 SVGPVVIASSGAGTPDHFSEVFEEDKRICRACCRHPPPERGYQSQVKEHLOEBERIEVRI 593

RESULT 9

AA194241

ID AAY94241 standard; Protein; 593 AA.

XX AC AAY94241;

XX 08-AUG-2000 (first entry)

XX Cress mutant cyclase, K404A substitution.

XX Glutamine amidotransferase; histidine biosynthesis; cyclase;

XX mouse ear cress; site-directed mutagenesis; mutein; substitution;
 mutant.

XX Arabidopsis thaliana.

OS Synthetic.


```

176 IRVPHIGNALQVGDSEILDDVGNHRVYFVHSYRAIPSDENKDWISSTCNYSGESFISI 235
200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKRVIACLDVRAN 259
236 RRGNVHAVQHPKSGVGLSVLRFLHPKL--PATQKPMGKASKLAKRVIACLDVRTN 293
260 DNGDLVVTGKQDYDVRERTEENRNLGKPVLAGOYLLDGADEVSLNITGFRDPLGD 319
294 DKGDLVVTGKQDYDVRERTEENRNLGKPVLAGOYLLDGADEVSLNITGFRDPLGD 353
320 LPMQLVQLRASENVFVPLTVGGGIRFTDANGRYYSLEVASFYFRSGADKVSIGSDAVY 379
354 LPMQLVQLRQTSKNVFPVLTGCGIRFTDASGRYYSLEVAASYFRSGAKMSIGSDAVF 413
380 TABEYIKTGKTKSSIEQISTVYGNQAVVSTDPRRVYLKRPDEVEFKAIVKSHPGNG 439
414 AAEFIKSGVKTGKSSLEQISRVYGNQAVVSTDPRRVYVNHPPDVPYKIVRTNFGPNG 473
440 EEWYQCTVNGGREGRPICAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 499
474 EEWYQCTVNGGREGRPICAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 533
500 AVNIPVIASSGAGVADHFSVFNETNASALAAGIF-HRKEVPIKAVKEHLLKEGIEVRL 558
534 SVGIPVIASSGAGTDPHFSEVFEDKRICACCRHPPPERGYQSQSVKEHLQERIEVRI 593

RESULT 14
AAY94243
ID AAY94243 standard; Protein; 593 AA.
AC AAY94243;
XX
XX
DT 08-AUG-2000 (first entry)
DE Cress mutant cyclase, D447A substitution.
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis; mutein; substitution;
KW mutant.
XX
XX
OS Arabidopsis thaliana.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 447
FT /note= "Wild-type Asp substituted by Ala"
XX
XX
PN WO200028053-A2.
XX
XX
PD 18-MAY-2000.
XX
XX
XX 04-NOV-1999; 99WO-US25950.
XX
XX
XX 05-NOV-1998; 98US-0107275.
XX
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX
XX Allen SM, Huang LL, Falco SC, Rafaleski AJ;
XX
XX
XX WPI; 2000-376564/32.
XX
XX
XX Novel polynucleotides encoding plant glutamine amidotransferase
XX homologues, useful for producing transgenic plants and as probes or
XX primers.
XX
XX
XX Example 9; Page -: 53pp; English.
XX
XX
XX The present sequence is a mutant of the mouse ear cress HisHf enzyme,
XX which is involved in histidine biosynthesis. It was produced by
XX site-directed mutagenesis, which was carried out to find the active
XX sites within the enzyme, which is made up of a cyclase and a glutamine
```

```

CC amidotransferase. The mutation in this case is in the cyclase
CC domain. Determining the active sites is useful for the identification of
CC herbicides and fungicides which inhibit the enzyme. This residue was
CC shown to form part of the active site.
CC Note: The present sequence is not shown in the specification but is
CC derived from the Arabidopsis thaliana wild-type His HF sequence given in
CC the sequence listing (see AAY94231).
XX
XX Sequence 593 AA;
SQ
Query Match 74.8%; Score 2152; DB 21; Length 593;
Best Local Similarity 76.5%; Pred. No. 1.3e-186;
Matches 413; Conservative 60; Mismatches 63; Indels 4; Gaps 3;
QY 21 AGAGDGSVVTLLDYGAGNVRSVRNARTIGFDIKOVQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 ASSTDSVVTLDDYGAGNVRSIRNALRHGFSIKDQVTFDGLINADRLIFPGVGAFAPAM 115
QY 81 DVLIRKGLAEALCTYIQNDRPFLGICLGLQLLFESSENGPIQGLGLIFCRVGRFPSSNG 140
DB 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLLFDSSEQNGPVKGLGVIPIGVGRFPDASAG 175
QY 141 LRVPHIGHALDIKEGSAILLDDVGNHRVYFVHSYRA-NAEDNKEWISSTCNYSIGSDFIASI 199
DB 176 IRVPHIGNALQVGDSEILDDVGNHRVYFVHSYRAIPSDENKDWISSTCNYSGESFISI 235
QY 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKRVIACLDVRAN 259
DB 236 RRGNVHAVQHPKSGVGLSVLRFLHPKL--PATQKPMGKASKLAKRVIACLDVRTN 293
QY 260 DNGDLVVTGKQDYDVRERTEENRNLGKPVLAGOYLLDGADEVSLNITGFRDPLGD 319
DB 294 DKGDLVVTGKQDYDVRERTEENRNLGKPVLAGOYLLDGADEVSLNITGFRDPLGD 353
QY 320 LPMQLVQLRASENVFVPLTVGGGIRFTDANGRYYSLEVASFYFRSGADKVSIGSDAVY 379
DB 354 LPMQLVQLRQTSKNVFPVLTGCGIRFTDASGRYYSLEVAASYFRSGAKMSIGSDAVF 413
QY 380 TABEYIKTGKTKSSIEQISTVYGNQAVVSTDPRRVYLKRPDEVEFKAIVKSHPGNG 439
DB 414 AAEFIKSGVKTGKSSLEQISRVYGNQAVVSTDPRRVYVNHPPDVPYKIVRTNFGPNG 473
QY 440 EEWYQCTVNGGREGRPICAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 499
DB 474 EEWYQCTVNGGREGRPICAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 533
QY 500 AVNIPVIASSGAGVADHFSVFNETNASALAAGIF-HRKEVPIKAVKEHLLKEGIEVRL 558
DB 534 SVGIPVIASSGAGTDPHFSEVFEDKRICACCRHPPPERGYQSQSVKEHLQERIEVRI 593
```

```

RESULT 15
AAY94245
ID AAY94245 standard; Protein; 593 AA.
XX
XX
XX AAY94245;
XX
XX
XX 08-AUG-2000 (first entry)
XX
XX
XX Cress mutant cyclase, D517A substitution.
XX
XX
XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
XX mouse ear cress; site-directed mutagenesis; mutein; substitution;
XX mutant.
XX
XX
XX Arabidopsis thaliana.
XX
XX
XX Synthetic.
XX
XX
XX Key Location/Qualifiers
XX
XX
XX Misc-difference 517
XX /note= "Wild type Asp substituted by Ala"
XX
XX
XX WO200028053-A2.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:25:12 ; Search time 22 Seconds
(without alignments)
1073.157 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877
Sequence: 1 HEKELASTKQNGFRIRAL.....EVPKAVKEHLLKEGIEVRL 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: Issued Patents AA.*
 - 2: /cgn2_6/ptodata/1/iaa/5A COMB.pdp.*
 - 3: /cgn2_6/ptodata/1/iaa/5B COMB.pdp.*
 - 4: /cgn2_6/ptodata/1/iaa/6A COMB.pdp.*
 - 5: /cgn2_6/ptodata/1/iaa/6B COMB.pdp.*
 - 6: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pdp.*
 - 7: /cgn2_6/ptodata/1/iaa/backfiles.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	454	15.8	311	4	US-09-352-991A-23527
2	442	15.4	288	4	US-09-328-352-6187
3	430.5	15.0	258	4	US-09-134-001C-4647
4	365	12.7	207	4	US-09-328-352-6054
5	357	12.4	192	3	US-08-953-139-2
6	357	12.4	192	3	US-08-953-139-4
7	355	12.3	182	4	US-09-134-001C-4652
8	351	12.2	371	4	US-09-352-991A-23484
9	350	12.2	235	4	US-09-352-991A-28794
10	268	9.3	265	4	US-09-352-991A-28727
11	264	9.2	251	2	US-08-846-762-11
12	184	6.4	118	2	US-08-846-762-10
13	138	4.8	261	4	US-09-352-991A-23486
14	129	4.5	265	4	US-09-328-352-6115
15	126.5	4.4	243	4	US-09-134-001C-4654
16	115	4.0	525	4	US-09-328-352-4897
17	108.5	3.8	881	3	US-08-960-780-32
18	108.5	3.8	881	3	US-09-073-898-32
19	107	3.7	188	4	US-08-858-207A-445
20	107	3.7	393	3	US-09-629-616-2
21	107	3.7	667	1	US-08-471-033-7
22	107	3.7	667	2	US-08-471-044-7
23	107	3.7	667	2	US-08-463-483A-7
24	107	3.7	667	2	US-08-471-046A-7
25	107	3.7	667	2	US-08-470-566B-7
26	107	3.7	667	2	US-08-469-334-7
27	107	3.7	667	3	US-09-300-529-7

28	107	3.7	852	1	US-08-471-033-36	Sequence 36, Appl
29	107	3.7	852	2	US-08-471-044-36	Sequence 36, Appl
30	107	3.7	852	2	US-08-463-483A-36	Sequence 36, Appl
31	107	3.7	852	2	US-08-471-046A-36	Sequence 36, Appl
32	107	3.7	852	2	US-08-470-566B-36	Sequence 36, Appl
33	107	3.7	852	2	US-08-469-334-36	Sequence 36, Appl
34	107	3.7	852	3	US-09-300-529-36	Sequence 36, Appl
35	107	3.7	884	1	US-08-471-033-5	Sequence 5, Appl
36	107	3.7	884	2	US-08-463-483A-5	Sequence 5, Appl
37	107	3.7	884	2	US-08-471-044-5	Sequence 5, Appl
38	107	3.7	884	2	US-08-471-046A-5	Sequence 5, Appl
39	107	3.7	884	2	US-08-470-566B-5	Sequence 5, Appl
40	107	3.7	884	2	US-08-469-334-5	Sequence 5, Appl
41	107	3.7	884	3	US-09-300-529-5	Sequence 5, Appl
42	107	3.7	1094	4	US-09-268-347-32	Sequence 32, Appl
43	107	3.7	1338	1	US-08-471-033-50	Sequence 50, Appl
44	107	3.7	1338	2	US-08-471-044-50	Sequence 50, Appl
45	107	3.7	1338	2	US-08-463-483A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-09-252-991A-23527
; Sequence 23527, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23527
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23527

Query Match 15.8%; Score 454; DB 4; Length 311;
Best Local Similarity 37.5%; Pred. No. 4.2e-38;
Matches 117; Conservative 45; Mismatches 92; Indels 58; Gaps 8;

QY	246	LAKRVTACLDVRANDNDLVVTGKQDYDVRTEENEVRNLGKPVLAGQYLLDGADEVS	305
DB	58	LAKRIIPCLDV---DNGRVV--KGVKFE-----NIRDAGDPVEIARRYDEQGADEIT	104
QY	306	FLNITGFRFPDGLPMLQVLRASENVVPLTVGGGIRDFDTNANGRYYSLEVASEYFR	365
DB	105	FLDITASVD---GRDTTLTHTVERMASQVPIPLTVGGGVSQVDIR-----NLN	150
QY	366	SGADKVSIGSDAVYTAEEYIKTVKTKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEV	425
DB	151	AGADKVSINTAAVFNPDE-----FVGEAADRFQSGCIVVAID-----	186
QY	426	EPKAIKVSHPGNGEYAWYQCTVNGRGREGRPTGAYELAKAVEELGABILLNCIDCDGQ	485
DB	187	---AKKVSAPG---EAPRWEIFTHGRKPTGLDAVLWAKWEDLGAGILLTSMDDGV	239
QY	486	KGKFDIDLKLSDAVNIPIVIASSGAGVADHSEVFNENETNSAALAAGTFRKEVPIKAV	545
DB	240	KSGYDLGVTRAISEAVNVPVIASGVGNLEHLAAGILEGKADAVLAASIFHFGEYTVPEA	299
QY	546	KEHLLKEGIEVR	557
DB	300	KAYLASRGIVVR	311

RESULT 2
US-09-328-352-6187
; Sequence 6187, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6187
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6187

Query Match 15.4%; Score 442; DB 4; Length 268;
Best Local Similarity 35.0%; Pred. No. 5.7e-37;
Matches 114; Conservative 52; Mismatches 98; Indels 62; Gaps 8;

QY 233 NKQKPMNGKASIKLAKRVIACLDVRANDGDLVVTGKQDYDVRERTEENVRNLGKPVVEL 292
DB 5 NFQRRRLFLKXKMLAKRIICLDV--DGRVV--KGQF-----LDIRAGDPVEV 51
QY 293 AGQYLDGADGVFLNTITGRDPLPMLQVLRASENVFVPLTVGGGIRDTFDTANGR 352
DB 52 ARRYNEQCADEITFLDITATHH---GRDTTYRTVERMAETVFVPLTVGGGVRKVEDIRA- 107
QY 353 YSSLEVASFYFSGADKVGSDAVYTAEEYIKTVGKTKSSIEQISTVYGNQAVVSI 412
DB 108 -----LLNAGADKVSINSAAVFNPE-----FVQEASQHFQAQCIIVAI 145
QY 413 DPRRVLYRKDVEVEFKAIKVSHPGNGEYAWYQCTVNGGREGRPICAYELAKAVEELGAGBILNCIDCDGQ 472
DB 146 DAKT-----GDNK-----WEIFTHGGRKPTGIDAIENAVKMDYGA 182
QY 473 GEILLNCIDCDGQKGFIDILKISDAVNIPVIASSGAGVADHFSEVFNETNASALAA 532
DB 183 GELLITSMADAGTKAGYDIALMRAINDRVITPTIASSGGVGNLQHLADGILQGGADAVLAA 242
QY 533 GIFHRKEVPKAVKHEHLKKEGIEVRL 558
DB 243 SIFHFGQYTIPEAKOYLAEQGIEMRL 268

RESULT 3
US-09-134-001C-4647
; Sequence 4647, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4647
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4647

Query Match 15.0%; Score 430.5; DB 4; Length 258;
Best Local Similarity 34.0%; Pred. No. 8.1e-36;
Matches 106; Conservative 53; Mismatches 92; Indels 61; Gaps 7;

QY 246 LAKRVIACLDVRANDGDLVVTGKQDYDVRERTEENVRNLGKPVVELAQYVLDGADEVS 305
DB 8 IKRVIPLCLDVK-----DGRVVGKIQF-----QSLRDIGNPVDLALYNEAGADELV 54
QY 306 FLNITGFRDPLPMLQVLRASENVFVPLTVGGGIRDTFDTANGRYYSSEVASEYPR 365
DB 55 FLDIS---KTEAGHDLMIIEVATAKQLFIPLTVGGGIGNLDD-----ITQLLN 100
QY 366 SGADKVSIGSDAVYTAEEYIKTVGKTKSSIEQISTVYGNQAVVSIIDPRRVLYRKPDV 425
DB 101 HGADKISLNSALKHPE-----LIRQASEKFGRCICIAID----- 136
QY 426 EFKAIKVSHPGNGEYAWYQCTVNGGREGRPICAYELAKAVEELGAGBILNCIDCDGQ 485
DB 137 -----SFYDKDREDYF---CTTHGGKKLTDVSVYDWTQVEVHGLGAGELLITSMHDDGM 186
QY 486 GKGFIDILKISDAVNIPVIASSGAGVADHFSEVFNETNASALAAAGIFHRKEVPKAV 545
DB 187 KQGFIDIEHLAKIKQLWNIPIIASGGGNAQHVFELFQOTDVSAGLAASILHDQETTVAEI 246
QY 546 KEHLKKEGIEVR 557
DB 247 KDKMREGGILVR 258

RESULT 4
US-09-328-352-6054
; Sequence 6054, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6054
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6054

Query Match 12.7%; Score 365; DB 4; Length 207;
Best Local Similarity 37.3%; Pred. No. 3.1e-29;
Matches 78; Conservative 35; Mismatches 74; Indels 22; Gaps 5;

QY 29 VTLLDYGAGNVRSVRNAIRTLGFDIKDVOKPEDILNAKRLIFPGVGAFAFAMDLIRKGL 88
DB 6 IALLDYGMLHSAAKALEYVGATVDVTDPKLIAQADKIVFPVGAMRDCMQGREAGI 65
QY 89 AEALCTYIONDRPFLGICLGLQLLFSSSENGPIQGLGILIPGRVGRFESSNGLRVPHIGW 148
DB 66 DEVVRKAFFN-KPVALICVGMQALLQSSSENGGVDALGIFDGIIVKHFPMQMEGLKVPBMGW 124
QY 149 ---HALD-----IKESGAILDDVGNQHVYVHSYRANAEDNKEMWISTCSVGDDFIA 197
DB 125 NQVHQMDPSHPMMWNIEQ-----DARFYFVHSYVEPKD-ENLVAATCEYGVNFT 174
QY 198 SIOKGNVHVAQVHFPEKSGGVGLSILRRFL 226
DB 175 AIHKDNLFAQVHFPEKSHTAGLQLLNKFV 203

RESULT 5
US-08-953-139-2
; Sequence 2, Application US/08953139
; Patent No. 6258578
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K.
; APPLICANT: Lonetto, Michael A.
; APPLICANT: Warren, Patrick V.


```
; TITLE OF INVENTION: NOVEL HIS5
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10056-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-953-139-2

Query Match 12.4%; Score 357; DB 3; Length 192;
Best Local Similarity 38.8%; Pred. No. 1.8e-28;
Matches 80; Conservative 35; Mismatches 77; Indels 14; Gaps 5;

QY 28 VVTLLDYGAGNVSVRNAITLGFIDKDVQKPEDILNAKELIFPGVGAFAPAMDVLIRKG 87
Db 1 MIVIDYGLGNISNVRKRAIEHLGYEVVSVNTSKIIDQAEITILPGVGHFADAMSEIKRLN 60
QY 88 LAEALCTYIQNDRPFGICLGLQLFPESSENGPIQGLIPGRVGRFESSNGLRVPHIG 147
Db 61 LNAILAK--NTDKMIGICLGMQMYEHSDE-GDASGLGFIPGNISRIQTE--YVPVHLG 115
QY 148 WHALDIKESAILDDVGNQHVYFVHSYRANAENKEMWISSTCSYGDPDFIASIQKGNHAV 207
Db 116 WNNLVSKH-----PMLNQDVYFVHSYQAPMSEN---VIAVAYGADIPAIVQFNYYIGI 166
QY 208 QFHPEKSGVGLSILRRFLNADSFNN 233
Db 167 QFHPEKSGTYGLQLRQAIQGGFIND 192

RESULT 6
US-08-953-139-4
; Sequence 4, Application US/08953139
; Patent No. 6258578
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K.
; APPLICANT: Lonetto, Michael A.
; APPLICANT: Warren, Patrick V.
; TITLE OF INVENTION: NOVEL HIS5
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
```

```
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10056-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-953-139-4

Query Match 12.4%; Score 357; DB 3; Length 192;
Best Local Similarity 38.8%; Pred. No. 1.8e-28;
Matches 80; Conservative 35; Mismatches 77; Indels 14; Gaps 5;

QY 28 VVTLLDYGAGNVSVRNAITLGFIDKDVQKPEDILNAKELIFPGVGAFAPAMDVLIRKG 87
Db 1 MIVIDYGLGNISNVRKRAIEHLGYEVVSVNTSKIIDQAEITILPGVGHFADAMSEIKRLN 60
QY 88 LAEALCTYIQNDRPFGICLGLQLFPESSENGPIQGLIPGRVGRFESSNGLRVPHIG 147
Db 61 LNAILAK--NTDKMIGICLGMQMYEHSDE-GDASGLGFIPGNISRIQTE--YVPVHLG 115
QY 148 WHALDIKESAILDDVGNQHVYFVHSYRANAENKEMWISSTCSYGDPDFIASIQKGNHAV 207
Db 116 WNNLVSKH-----PMLNQDVYFVHSYQAPMSEN---VIAVAYGADIPAIVQFNYYIGI 166
QY 208 QFHPEKSGVGLSILRRFLNADSFNN 233
Db 167 QFHPEKSGTYGLQLRQAIQGGFIND 192

RESULT 7
US-09-134-001C-4652
; Sequence 4652, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4652
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4652
```


Db 150 -----WGHVSFSENGTRDMKRSPL-----EWAQALEAGVG 181
QY 474 EILLNCIDCGQKGPDIKILISDAVNIPVIASSGADVHFSEVFNETNASALAAG 533
Db 182 EIFLNSIDRGVQKGFNDALVENIASNVHVPVIACGAGSIADLIDLFERTCVS-AVAG 240
QY 534 ---IFHRK 538
Db 241 SLFVFHGK 248

RESULT 11
US-08-846-762-11
; Sequence 11, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-11

Query Match 9.2%; Score 264; DB 2; Length 251;
Best Local Similarity 30.8%; Pred. No. 1.1e-18;
Matches 92; Conservative 44; Mismatches 87; Indels 76; Gaps 12;
QY 248 KRVIACLDVRANDGDLVVTGQDYDVRTEENVRNLGKPVLAGQYYLDGAEVSPL 307
Db 4 RRVIPCLLLK--DRG-LVKT-----VKPEKPYGDPINATRIFFEKEVDELILL 50
QY 308 NITGFRDPLGDLPLQVLQASENVFVPLTVGGGIRDFTDANGRYYSLEVASFYRSFG 367
Db 51 DIDASR---LNQEPNYELIAEVAGCEPICYGGIK-----TLEHAETKIFSLG 96
QY 368 ADKVSIGSDAVYTAEEYIKTKGKSSIEQISTVYGNQAVVVSIDPRVYLKRPDEVEF 427
Db 97 VEKVSINTAALMDL-----SLIRRIADKFGSQSVGSDICRKGf-----135
QY 428 KAIKVSHPGNGEYAVYQCTV---NGREGR--PIGAYELAKAVEELGAGEILLNCIDC 482
Db 136 -----WGHVSFSENGTRDMKRSPL-----EWAQALEAGVGEIFLNSIDR 176
QY 483 DGQCKGPDILIKLISDAVNIPVIASSGADVHFSEVFNETNASALAAG---IFHRK 538
Db 177 DGQCKGPDNALVENIASNVHVPVIACGAGSIADLIDLFERTCVS-AVAGSLFVFHGK 234

RESULT 12
US-08-846-762-10
; Sequence 10, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30

; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-10

Query Match 6.4%; Score 184; DB 2; Length 118;
Best Local Similarity 36.4%; Pred. No. 5e-11;
Matches 44; Conservative 19; Mismatches 50; Indels 8; Gaps 3;
QY 107 LGLQLLFESSENGPIQGLGLIPRGVRFPSSNGLRVPHIGHALDIKESAILDDVQNG 166
Db 2 LGLR-----SSEGAEPGLGWDMSVRFERRDRKVPHGMQNVSPQLEHPILSGINEQ 55
QY 167 -HVYFVHSYRANAENKEMWISSTCSYGDPIASIQKGNVHAVQFHPKSGGVCLSLRRF 225
Db 56 SRPYFVHSYTMVPKOPDD--ILLSCNYGQKFTAAVARDNVFGFQFHPKSHKFGMLKKNF 114
QY 226 L 226
Db 115 V 115

RESULT 13
US-09-252-991A-23486
; Sequence 23486, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23486
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23486

Query Match 4.8%; Score 138; DB 4; Length 261;
Best Local Similarity 21.4%; Pred. No. 1.1e-05;
Matches 56; Conservative 38; Mismatches 90; Indels 78; Gaps 8;
QY 252 ACIDVRANDGDLVVTGQDYDVRTEENVRNLGKPVLAGQYYLDGAEVSFLNITG 311
Db 29 ACVRLRQGLMEDATVFSDD-----PVMAAKWVDGGRRLHLVDLNG 70
QY 312 -FRDFPLGDLPLQVLQASENVFVPLTVGGGIRDFTDANGRYYSLEVASFYRSFGADK 370
Db 71 APECKVNGEVVTAIRYPD---LPIQGGIR-----SLEITIEHVRAVSYS 116
QY 371 VSIGSDAVYTAEEYIKTKGKSSIEQISTVYGNQAVVVSIDPRVYLKRPDEVEFK 428
Db 117 VIIGTKAV-KQPEFVGEACRAPGK-----VIVGLDAKQGFV-----152
QY 429 AIKVSHPGNGEYAVYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCGGKG 488
Db 153 -----ATDGAESVQVVIDLARRFEADGVSAIVYTDISKGMQMG 193
QY 489 FDIDLIKLSDAVNIPVIASSG 510
Db 194 CNVQATAALANATRIPIVASSG 215

RESULT 14

[illegible]

RESULT 15

```

US-09-134-001C-4654
; Sequence 4654, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4654
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4654

```

```

      : :: :   !: | ||||| :       :   :   | | | | :   :
78  RSLRLTKYK---PIEVGGIR-----SKQTIENYHSGIDYCVGT----- 115

      : : : : : : : : : : : : : : : : : : : : : : : : : :
384 YIKTGVKTGKSISBQISTVYGNOAVVSIDPRRVYLKKPDEVFEKAIKVSHPGPN---- 438
      : : : : : : : : : : : : : : : : : : : : : : : : : :
116 -----KGIOIEWLT-----HMTHQFPNKLKLYLS 138

      : : : : : : : : : : : : : : : : : : : : : : : : : :
439 ---GEEVAWQCTVNGREGRPICAYELAKAVEELGAGELLNCIDCGOGKGPDI LI 494
      : : : : : : : : : : : : : : : : : : : : : : : : : :
139 VDAFGE-----KIKINGKEKAKUNLPDYVAKIEHLPLGGVIYTDISKCKLSPFDLT 193

      : : : : : : : : : : : : : : : : : : : : : : : : : :
495 KLISDAVNIPVIASGAGVADHFSEVF--NETNASAALAAGI FHRKE 539

      : : : : : : : : : : : : : : : : : : : : : : : : : :
194 GRILAYTSLPVIASGG---IRHOEDLFRELSNVHAIVGKAAHLDE 237

```

Search completed: January 20, 2004, 16:28:40
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:26:42 ; Search time 39 Seconds
(without alignments)
2925.654 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877

Sequence: 1 HEKELASTKPGQFRIRAAAL.....EVPKAVKHEHLKEGIEVRL 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1397	48.6	576	12	US-10-369-493-3441
2	1363	47.4	552	12	US-10-369-493-1500
3	1327.5	46.1	502	12	US-10-369-493-22552
4	667.5	23.2	343	12	US-10-369-493-13241
5	529.5	18.4	271	12	US-10-369-493-987
6	522	18.1	301	12	US-10-369-493-1220
7	510.5	17.4	438	12	US-10-369-493-20868
8	499.5	17.4	272	12	US-10-369-493-21460
9	486.5	16.9	273	12	US-10-369-493-11321
10	481.5	16.7	258	10	US-09-738-626-5791
11	478.5	16.6	273	12	US-10-369-493-18588
12	472	16.4	256	12	US-10-369-493-8203
13	465	16.2	256	12	US-10-369-493-13786
14	457	15.9	251	12	US-10-369-493-16690
15	452.5	15.7	256	12	US-10-369-493-19698

16	451	15.7	257	12	US-10-369-493-4591	Sequence 4591, Ap
17	451	15.7	257	12	US-10-369-493-7348	Sequence 7348, Ap
18	449.5	15.6	259	12	US-10-369-493-20736	Sequence 20736, A
19	448	15.6	253	12	US-10-369-493-2990	Sequence 2990, Ap
20	443.5	15.4	255	12	US-10-369-493-19005	Sequence 19005, A
21	438.5	15.2	251	15	US-10-156-761-13687	Sequence 13687, A
22	436	15.2	259	12	US-10-369-493-9689	Sequence 9689, Ap
23	432.5	15.0	252	12	US-10-369-493-23318	Sequence 23318, A
24	430.5	15.0	255	12	US-10-369-493-20110	Sequence 20110, A
25	427	14.8	255	12	US-10-369-493-8759	Sequence 8759, Ap
26	423.5	14.7	252	12	US-10-369-493-17439	Sequence 17439, A
27	419	14.6	260	12	US-10-369-493-9934	Sequence 9934, Ap
28	416.5	14.5	250	12	US-10-369-493-10338	Sequence 10338, A
29	410.5	14.3	261	12	US-10-369-493-2739	Sequence 2739, Ap
30	406	14.1	263	12	US-10-369-493-12104	Sequence 12104, A
31	404	14.0	254	12	US-10-369-493-20963	Sequence 20963, A
32	399.5	13.9	259	12	US-10-369-493-18411	Sequence 18411, A
33	391.5	13.6	260	12	US-10-369-493-17089	Sequence 17089, A
34	390	13.6	259	12	US-10-369-493-9026	Sequence 9026, Ap
35	389	13.5	265	12	US-10-369-493-531	Sequence 531, App
36	387	13.5	260	12	US-10-369-493-11468	Sequence 11468, A
37	387	13.5	260	12	US-10-369-493-14332	Sequence 14332, A
38	387	13.5	260	12	US-10-369-493-14650	Sequence 14650, A
39	386	13.4	264	12	US-10-369-493-10789	Sequence 10789, A
40	385	13.4	250	12	US-10-369-493-10990	Sequence 10990, A
41	383	13.3	254	12	US-10-369-493-15124	Sequence 15124, A
42	375.5	13.1	252	12	US-10-369-493-7967	Sequence 7967, Ap
43	373.5	13.0	258	12	US-10-369-493-21256	Sequence 21256, A
44	361.5	12.6	258	12	US-10-369-493-883	Sequence 883, App
45	354	12.3	255	12	US-10-369-493-15682	Sequence 15682, A

ALIGNMENTS

RESULT 1

US-10-369-493-3441
; Sequence 3441, Application US/10369493
; Publication No. US20030233675A1

GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 3441

; LENGTH: 576

; TYPE: PRT

; ORGANISM: Neurospora crassa

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(576)

; OTHER INFORMATION: unsure at all xaa locations

US-10-369-493-3441

Query Match 48.6%; Score 1397; DB 12; Length 576;

Best Local Similarity 51.0%; Pred. No. 1.5e-126;

Matches 290; Conservative 81; Mismatches 156; Indels 42; Gaps 9;

Qy 29 VTLLDYAGNVSRRNAIRTLGFDIKDVKPDLINA----- 65

Db 4 VHLDDYAGNVSRRNAIRTVGVSVWRSPEVANADVCANLSTETPLTRTCVSHV 63

Qy 66 KRLIFPGVGAFAPMVDLIRKGLAEALCTYIQNDRPFLGICLGLQLLFFSEENSGPIQGL 125

Db 64 XKLIPGVGHGHCLSQASAGYMAPIRAHIESGKPFMGICVGLQALFEGSSDPNCPGL 123
Qy 126 GLIPGRVGRFESSNGLRVPHIGHALDIKESAILDDVGNQHVYFVHSYR-----ANAEDN 181
Db 124 GVLPGKLDREDDTS-KSVPHIGMNDASCPNPNLUGLNPDSKYVYHSHYKMPYTKGQLES 182
Qy 182 KEWISSTCSY-GDDFIASIOKGNHVAQVPHPEKSGGVLSTLRFLN-----ADSFNNKR 235
Db 183 QGNAVATGVYGETFICGAIAGNVMATQPHPEKSGVAGLVRALFADGSGAALASHPPQ 242
Qy 236 QKPMNGK---ASK--LAKRVIACLDVRANDNGDLVVTGQDQYDVRETEENEVRNKGVP 290
Db 243 EVPNAGDVALYSKEGLTRRVIACLDVRTNDQGLVVTGQDQYDVREKSDDRNVRNKGVP 302
Qy 291 ELAQVYLDGADSVFNITGRDPLGLDPLQVLRASENVFVPLTVGGGIRDFDAN 350
Db 303 EMARKYEQGADEVTFNLNITSFRDPCVADLPMLBILRLTSTKTVFVPLTVGGGIRDFD 362
Qy 351 GRYSSELEAVASEYFRSGADKVSIGSDAVYTAEEYIKTGVT--GKSSIEQISTVYGNQAV 408
Db 363 GTKVSALEIATMYFQSGADKVSIGSDAVIAEEYASG-KTLFGNTAIEQISKAYGNQAV 421
Qy 409 VVSDPRVYLRKPEVEPKAKVSHPGPNGEYAMVYQCTVNGGREGRPIGAYELAKAVE 468
Db 422 VVSDPRVYLRKPEVEPKAKVSHPGPNGEYAMVYQCTVNGGREGRPIGAYELAKAVE 481
Qy 469 ELGAGEILLNCIDCGOGKGFDDILIKLSDAVNIPVIASSGAGVADHSEVFNETNASE 528
Db 482 AMGCGEILLNCIDKDGNSGFDLEIRQVAAVRIPVIASSGAGNPGHFEVFRFTTDA 541
Qy 529 ALAAGIFHRKEVPIKAVKEHLKEGIEVR 557
Db 542 ALGAGMFRGEYTVQVQKEELKARGLVVR 570

RESULT 2
US-10-369-493-1500
; Sequence 1500, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1500
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1500

Query Match 47.4%; Score 1363; DB 12; Length 552;
Best Local Similarity 50.4%; Pred. No. 2,9e-123;
Matches 281; Conservative 91; Mismatches 149; Indels 36; Gaps 11;
Qy 28 VVTLDDYGAGNVRVRNAITLFGDIDKVDKPED--ILNAKRLIPPGVAFAPAMDVLR 85
Db 3 VVHVIDVESGNLQSLTAIEHLGVEQLVKSPPKDFNISGTSRLTLPVGNYGHFVDNLFN 62
Qy 86 KGLAEALCTYQNDPRFLGICLGLQLLPESSENGPIQGLIPGRVGRFESSNGLRVPH 145
Db 63 RGEKPIREYESGKPIKMGICVGLQALFAGSVESPKSTGLNYIDFKLSRFDDBS-KVPE 121
Qy 146 IGMHALDIKESGAI-LDDVGNQHVYFVHSYRA--NAE-----DNKEWISSTCSYG-DDFI 196

Db 122 IGMNSCIPSENLPFGLDPY--KRYVFGHFAAILNSEKKKQNLNDGDKTAKAKYGSSEFI 179
Qy 197 ASIOKGNHVAQVPHPEKSGGVLSTLRFLNADSFNNKRQKP-----MNGKA 243
Db 180 AAVNKNITATQPHPEKSGKAGLVNENFL-----KQSPPIPNYSABEKLMDYDYS 232
Qy 244 S-KLAKRVIACLDVRANDNGDLVVTGQDQYDVRETEENEVRNKGVPVLAGOYVLDGAD 302
Db 233 NYGLTRRIIACLDVRTNDQGLVVTGQDQYDVREKSDGKVRNKGVPVLAQKYVQGGAD 292
Qy 303 EYVFLNITGRDPLGLDPLQVLRASENVFVPLTVGGGIRDFDANGRYYSSELEVASE 362
Db 293 EVTFNLNITSFRDPCVADLPMLBILRLTSTKTVFVPLTVGGGIRDFDANGRYYSSELEVASE 352
Qy 363 YFRSGADKVSIGSDAVYTAEEYIKTGVT--GKSSIEQISTVYGNQAVVSDPRVYLRK 421
Db 353 YFRSGADKVSIGSDAVYTAEEYIKTGVT--GKSSIEQISTVYGNQAVVSDPRVYLRK 412
Qy 422 PDEVEFKAIKVSHPGPNGEYAMVYQCTVNGGREGRPIGAYELAKAVEELGAGELLNCID 481
Db 413 QADTKNKVETETEPGNGEKYCWYQCTIKGGRSRLDGVWELTRACEALGAGELLNCID 472
Qy 482 CDGQGGFDILIKLSDAVNIPVIASSGAGVADHSEVFNETNASEAALAAAGIFHRKEVP 541
Db 473 KDGNSGYDLELIEHVKDAVKIPVIASSGAGVPEHFEAFKTRADACLGAGMFRHGFT 532
Qy 542 IKAVKEHLKEGIEVR 558
Db 533 VNDVKBYLLEHGLKVRM 549

RESULT 3
US-10-369-493-22552
; Sequence 22552, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22552
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22552

Query Match 46.1%; Score 1327.5; DB 12; Length 502;
Best Local Similarity 52.4%; Pred. No. 6.9e-120;
Matches 268; Conservative 82; Mismatches 128; Indels 33; Gaps 8;
Qy 68 LIPFGVGAFAPAMDVLRKGLAEALCTYQNDPRFLGICLGLQLLPESSENGPIQGL 127
Db 3 LIPFGVGNFGVCDLSLAKQGLFLEPLRRYALSGKPFMAVCVGIQALFEGSVAPHSKGLGV 62
Qy 128 IPGRVGRFESSNGLRVPHIGHALDIKESAILDDVGNQ-----HVTYFVHSYRANAE 179
Db 63 FPLGVQRFNDND-KTVPHIGWNS-----CAVASDTSKEFFGMRRPHDKFYFVHSYMIPEK 115
Qy 180 D---NKEWISSTCSYGDD-FIASIOKGNHVAQVPHPEKSGGVLSTLRFLNADSFNNKR 235
Db 116 GLILPPEFKIATTKYGNFVGAIVKQNFATQPHPEKSGSAGRLCKAFLTGN-----Y 170
Qy 236 QKPMNGKASK-----LAKRVIACLDVRANDNGDLVVTGQDQYDVRETEENEVRNKG 287


```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1220
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
; US-10-369-493-1220

Query Match      18.1%; Score 522; DB 12; Length 301;
Best Local Similarity 40.6%; Pred. No. 5.8e-42;
Matches 127; Conservative 44; Mismatches 102; Indels 40; Gaps 7;

QY 246 LAKRVIACTLDVRANDGDLVVTGDDQYDVRERTEENVRNLGKPVLAGQYVLDGADDEV 305
DB 29 LAKRIIPCLDC-----DLQVPG---RVVKGVEFKQIRVAGDPVELATRYEDGADIV 79

QY 306 FLAITGRDFPLGDLPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYVYSSLEVAEYFR 365
DB 80 FLOITASHE---RRETMTHTVIEATTENVFVPCVGGGIRKPED-----YFKWL-----K 125

QY 366 SGADKVSIGSDAVVTAEEYIKTVGKSKSIEQISTVYGNQAVVVSIDPRRVYLKRPDEV 425
DB 126 AGADKCSNTAANKPE-----LINEASDLVGSQACVVAIDAKERYIENPRES 173

QY 426 EFKAIKVSHPGNGEYAVYQCTVNGGREGPIGAYELAKAVELGAGEILLNCIDCDGQ 485
DB 174 DERFIEVDG-----YCWYECISYGGREFTGIDAVKAMECDQDQAGEILLTSMRDGT 228

QY 486 GKGFDDILKISDANVPIVASSGAGVADHSEFVNETWASALAAGIHFHKEVPIKAV 545
DB 229 KMGYDIPFLTMSGNDIPVASSGGEPEHIEAFTDGDADAALAASIFHFNEYVPVAV 288

QY 546 KEHLKEGIEVRL 558
DB 289 KEYLRSKGVPIRL 301

RESULT 7
; US-10-369-493-20868
; Sequence 20868, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20868
; LENGTH: 438
; TYPE: PRT
; ORGANISM: SYNECHOCOCCUS SP. WH 8102
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(438)
; OTHER INFORMATION: unsure at all Xaa locations
; US-10-369-493-20868

Query Match      17.7%; Score 510.5; DB 12; Length 438;
Best Local Similarity 29.9%; Pred. No. 1.4e-40;
Matches 161; Conservative 79; Mismatches 182; Indels 117; Gaps 19;

QY 27 SVWTLDDYAGNVRVSRNAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAMDVLIRK 86

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1220
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
; US-10-369-493-1220

Query Match      17.4%; Score 499.5; DB 12; Length 272;
Best Local Similarity 40.1%; Pred. No. 7.5e-40;
Matches 128; Conservative 47; Mismatches 89; Indels 55; Gaps 10;

QY 246 LAKRVIACTLDVRANDGDLVVTGDDQYDVRERTEENVRNLGKPVLAGQYVLDGADDEV 304
DB 2 LTKRIIPCLDIK-----DGRVVKG-----TKFLNLRDAGDPVELA-QYDDEGADEL 47

QY 305 SFNLITGF---RDFPLGDLPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYVYSSLEVAS 361
DB 48 VFLDITASAEKRDI-----IIDVERTAEKVFPLTVGGGIR-----SIEDFR 90

QY 362 EYFRSGADKVSIGSDAVVTAEEYIKTVGKSKSIEQISTVYGNQAVVVSIDPRRVYLK 421
```


Db 91 RILRAGADKVSINTAAV-----KNPLIKEASEIFGSCVVAIDAKRHYNE 138
QY 422 P--DEVEFRAIKVHPNGEEVAYQCTVNGREGRPIGAYELAKAVEELGAGEILLNC 479
Db 139 DEIDKINKNVKVE-----DGVCFEVIYGRKETGIDAINWAKKVELGAGEILLTS 192
QY 480 IDCQGGKGFIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNASAAAGIFHRKE 539
Db 193 IDKDGTKSGVDLILTKEISKSVKLPVIASGGCGKPEHVYEAFFYVGKADAALMAGILHYRE 252
QY 540 VPIKAVKEHLLKEGIEVRL 558
Db 253 YTIEEIKKYCADRGIPMLR 271

RESULT 9

US-10-369-493-11321
; Sequence 11321, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11321
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
US-10-369-493-11321

Query Match 16.9%; Score 486.5; DB 12; Length 273;
Best Local Similarity 37.8%; Pred. No. 1.4e-38;
Matches 119; Conservative 51; Mismatches 100; Indels 45; Gaps 7;
QY 246 LAKRVIACLDVRANDGDLVVTGQDYDVRTEENEVRNLGKPVLAGQYVLDGADSVS 305
Db 2 LTRKIIPCLDVTDRAAGCVKGVGFVDLKE-----AGDPVELAKRYNEEGADELV 52
QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
Db 53 FLDTITSAQ--GRKTMIDVIERTADEVFPLTIGGI-----NSIDAIRQILR 98
QY 366 SGADKVSIGSDAVYTAEBYIKTVGKTSIEQISTVYGNQAVVVSIDPRR--VYLKPD 423
Db 99 AGADKVSVNTSAVKNPD-----FIKSSDIFGAQCIIVTADCRNRDITKNPD 146
QY 424 EVEFKAIVSHPGNGEEVAYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCD 483
Db 147 KT---ILEEDGTP-----AWYEVYGGREATGIDAVQWARKAEELGSGEILLTSMRD 198
QY 484 GQGGKGFIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNASAAAGIFHRKEVP 543
Db 199 GTCAGYDLPITKLSBELDIPVIASGGVGNPQHIYEGFSIGKADAALAAASIFHREYSIK 258
QY 544 AVKEHLLKEGIEVRL 558
Db 259 EVKEYLRDRREIPVRL 273

RESULT 10

US-09-738-626-5791
; Sequence 5791, Application US/09738626
; Publication No. US20020197605A1

; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5791
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5791

Query Match 16.7%; Score 481.5; DB 10; Length 258;
Best Local Similarity 39.4%; Pred. No. 3.9e-38;
Matches 123; Conservative 49; Mismatches 83; Indels 57; Gaps 10;
QY 246 LAKRVIACLDVRANDGDLVVTGQDYDVRTEENEVRNLGKPVLAGQYVLDGADSVS 305
Db 3 VAIRVIPCLDV--DNGRVV--KGVNFE-----NLKADGPVELAKRYNEEGADELT 49
QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
Db 50 FLDTITASKH--GRGTMLDVVRRTADQVFIPLTVGGVR-----SEEDVDQLLR 95
QY 366 SGADKVSIGSDAVYTAEBYIKTVGKTSIEQISTVYGNQAVVVSIDPRRVLKRPDEV 425
Db 96 AGADKVSVNTSAIARPE-----LLSELSKRFCAQCVILSVDAARRV---PEG- 138
QY 426 EFKAIKVSHPGNGEEVAYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCDGQ 485
Db 139 -----GTPQPSG-----FEVTHGGSKSAELDAIEWAKRGEELGVGEILLNSMDGDT 186
QY 486 GKGFDIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNASAAAGIFHRKEVP 545
Db 187 KNGFDLELEKVRAAVSIPVIASGGAGKAEHFPFPAV-AAGANAVLAATIFHFREVTIAEV 245
QY 546 KEHLLKEGIEVR 557
Db 246 KGAIKDAGFEVR 257

RESULT 11

US-10-369-493-18588
; Sequence 18588, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 18588
;; LENGTH: 273
;; TYPE: PRT
;; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-18588

Query Match 16.6%; Score 478.5; DB 12; Length 273;
Best Local Similarity 36.4%; Pred. No. 8.3e-38;
Matches 116; Conservative 45; Mismatches 99; Indels 59; Gaps 8;
QY 246 LAKRVIACLDVRANDGDLVVTGKQDYDVRTEENEVNRLGKPVLAGOYLLDGADEVS 305
DB 3 LTRKVPICIDVDLDGEPAYVTVNFE-----ELATGDPVEMAKRYNAGADEFV 54
QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
DB 55 FLDTASAE---GRETMLDTVSAVADEVFPLTVGGGIRDTDIR-----ETLR 100
QY 366 SGADKVSIGSDAVYTAEEYIKTVGKTSIEIQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 101 AGADKVSINGAI-----ADPSLVDRGAKAFSGQCVISVDARRF-----DE- 143
QY 426 EFKAIVKSHPGNGEYAWYQCTVNGREGRPVIGAYELAKAVEELGAGETLLN 478
DB 144 -----QGHVTVQDVGSECFVTHGREGTGMDALEWQEAORRGAGELFVN 191
QY 479 CIDCGGKGFDDILKILSDAVNIPVIASSGAGVADHSEVFNENETNASAAALAGIFHRK 538
DB 192 SIDADGTQDGVDPVPLTAAVCDVSTPVVASSGCGAGDMADAV-DAGADAALAAISIFHD 250
QY 539 EVPIKAVKEHLKEGIEVR 557
DB 251 EYSIAETKETLADAGYPIR 269

RESULT 12
US-10-369-493-8203
;; Sequence 8203, Application US/10369493
;; Publication No. US20030233675A1
;; GENERAL INFORMATION:
;; APPLICANT: Cao, Yongwei
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Chen, Xianfeng
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493
;; PRIOR FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 8203
;; LENGTH: 256
;; TYPE: PRT
;; ORGANISM: Thermobifida fusca
US-10-369-493-8203

Query Match 16.4%; Score 472; DB 12; Length 256;
Best Local Similarity 39.1%; Pred. No. 3.2e-37;
Matches 122; Conservative 43; Mismatches 89; Indels 58; Gaps 10;
QY 246 LAKRVIACLDVRANDGDLVVTGKQDYDVRTEENEVNRLGKPVLAGOYLLDGADEVS 305
DB 3 LARVIFCLDV---DGRVY--KGVPF-----QNLRDAGDPVELARYDAEGADELT 49
QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
DB 50 FLDTASSS---NRETTYDVRRTAEQVFIPLTVGGGV-----STEDVDRLLR 95

QY 366 SGADKVSIGSDAVYTAEEYIKTVGKTSIEIQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 96 AGADKVSINTAAVRPE-----LIREIAQFGSQVLVSADVRRT-----V 136
QY 426 EFKAIVKSHPGNGEYAWYQCTVNGREGRPVIGAYELAKAVEELGAGETLLNCIDCGQ 485
DB 137 NGTAT-----PSG-----FEITTHGGRQGTGIDAVEMVQQABELGAGEILLNSMDADGT 185
QY 486 GKGFDDILKILSDAVNIPVIASSGAGVADHSEVFNENETNASAAALAGIFHRKEVPIKAV 545
DB 186 KSGFDLELIRAVRKAVNVPLIASGGAGAVEHPAPV-DAGANAVLAASVHFGEITISDV 244
QY 546 KEHLKEGIEVR 557
DB 245 KAELEKAGYVPR 256

RESULT 13
US-10-369-493-13786
;; Sequence 13786, Application US/10369493
;; Publication No. US20030233675A1
;; GENERAL INFORMATION:
;; APPLICANT: Cao, Yongwei
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Chen, Xianfeng
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493
;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 13786
;; LENGTH: 256
;; TYPE: PRT
;; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13786

Query Match 16.2%; Score 465; DB 12; Length 256;
Best Local Similarity 38.1%; Pred. No. 1.5e-36;
Matches 119; Conservative 46; Mismatches 89; Indels 58; Gaps 8;
QY 246 LAKRVIACLDVRANDGDLVVTGKQDYDVRTEENEVNRLGKPVLAGOYLLDGADEVS 305
DB 3 LAKRIIPCLDV---DNGRVY--KGVPF-----NIRDAGDPVEIARRYDEQGADEIT 49
QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
DB 50 FLDTASVD---GRDTLTHVVERMASQVFIPLTVGGGVTVQDIR-----NLIN 95
QY 366 SGADKVSIGSDAVYTAEEYIKTVGKTSIEIQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 96 AGADKVSINTAAVRPE-----FVGEAAQHFSGQCIVVAILD----- 131
QY 426 EFKAIVKSHPGNGEYAWYQCTVNGREGRPVIGAYELAKAVEELGAGETLLNCIDCGQ 485
DB 132 ---AKKVSFGF---ETPRWEIPTHGGRKPTGIDAVEMAKQWEGGLGAGEILLTSMDDQGM 184
QY 486 GKGFDDILKILSDAVNIPVIASSGAGVADHSEVFNENETNASAAALAGIFHRKEVPIKAV 545
DB 185 KNGFDLGVTFRALSDALGIPVIASSGGVGNLQHLADGILEGHASAVLAASVHFGEITYVQEA 244
QY 546 KEHLKEGIEVR 557
DB 245 KAYMSKRGIVMR 256

RESULT 14
US-10-369-493-16690

[illegible]

Qy	546	KEHLLKEGIEVR	557
		:	
Db	241	KEVIRQGGIAVR	252

Search completed: January 20, 2004, 16:28:05
Job time : 22 secs

RESULT 15
A99976
C:Accession: A99976
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
R:Kuroda, M.; Ohra, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Jui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaico, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A99976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-252 <KUR>
A:Cross-references: GB:BA000018; PID:gl3702631; PIDN:BA843771.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: hlsf
A:Superfamily: cyclass hisf

Query Match	15.7%	Score 450.5;	DB 2;	Length 252;
Best Local Similarity	35.9%;	Pred. No. 7.3e-24;		
Matches 112; Conservative 55; Mismatches 84; Indels 61; Gaps 8;				
Qy	246	LAKRVIACLDVRANDNGDLVYTKGQVDVRETEENEVRNLKGVELAGQYYLGDAREVS	305	
Dd		: : :	:	:
	2	IKKR IIPCLDVK-----DGRVYKGIQF-----KGLRDIGNPVDLAITYNEAGADELV	48	
Qy	306	FLNITGTFRPPLGLDPLMLQLORASENVFPLTVGGGIIRDFTDANGRYSSYSLAVESEYER	365	
Dd	49	FLDISKTET-----GHSLMDEVIEQTASRLFIPLTVGGGIQ-----SLDDITQLLN	94	
Qy	366	SGADKVSTIGSDAVYTABEYIKTGVKTKGSSTIEQISTVYGNQAVVSDPRRVYLKPDEV	425	
Dd	95	HGADKVSLSNSALKNPQ-----LIQSADKEGRCICIAID-----	130	
Qy	426	EFKAIKVSHPGNGEEYAWYOCTVNGREGRPICAYELAKAVEBLAGELIINCDCDQG	485	
Dd	131	-----SYYPD--ERRAHYCCT-HGGKKMTNIKYVDWVQQVEQLGAGELLVTSMGHDM	180	
Qy	486	GKGFDDIDLILISDAVINPVITASSAGVAHFSEVFNETNASALACIFHRKEVPYKAV	545	
Dd	181	KQGPDIHKLAKISKLVNIPITASGGGNAOCHFVELFNOTDVSAGLAASILHDRETTVVQSI	240	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:20:32 ; Search time 18 Seconds
(without alignments)

1457.828 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877

Sequence: 1 HEKELASTKPFQNGFRIRAL.....EVPKAVKEHLLKEGIEVRL 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2284.5	79.4	522	1 HIS6_ARATH	Q98230 arabidopsis
2	1431.5	49.8	541	1 HIS6_SCHPO	Q94303 schizosacch
3	1363	47.4	552	1 HIS6_YEAST	P33734 saccharomyc
4	1315.5	45.7	553	1 HIS6_EMENI	Q94099 emericella
5	529.5	18.4	271	1 HIS6_ARCFU	Q29439 archaeoglob
6	527	18.3	272	1 HIS6_METKA	Q8tyw8 methanopyru
7	522	18.1	274	1 HIS6_METH	Q27398 methanobact
8	510.5	17.7	253	1 HIS6_THETN	Q8r885 thermoaer
9	499.5	17.4	272	1 HIS6_METJA	Q57854 methanococc
10	496.5	17.3	253	1 HIS6_AQUAE	Q66567 aquifex aeo
11	493.5	17.2	273	1 HIS6_METAC	Q8tt96 methanosarc
12	489.5	17.0	273	1 HIS6_METMA	Q8pw32 methanosarc
13	481.5	16.7	253	1 HIS6_CLOAB	Q97kx8 clostridium
14	481.5	16.7	258	1 HIS6_CORGL	Q31139 corynebacte
15	480.5	16.7	251	1 HIS6_LISIN	Q92e88 listeria in
16	480	16.7	254	1 HIS6_THET	Q9rp44 thermoaer
17	478.5	16.6	273	1 HIS6_HALN1	Q9hr52 halobacteri
18	474	16.5	251	1 HIS6_PYRPU	P58800 pyrococcus
19	472	16.4	253	1 HIS6_PYRAE	Q8zy16 pyrobaculum
20	467.5	16.2	251	1 HIS6_LISMO	Q8y955 listeria mo
21	467.5	16.2	267	1 HIS6_MYCTU	Q53908 mycobacteri
22	459.5	16.0	249	1 HIS6_SULTO	Q97020 sulfolobus
23	454	15.8	256	1 HIS6_PSEAE	Q9hu44 pseudomonas
24	450.5	15.7	252	1 HIS6_STAAM	Q99qk8 staphylococ
25	449.5	15.6	251	1 HIS6_SULSO	Q33774 sulfolobus
26	448	15.6	253	1 HIS6_THEMA	Q9x0c6 thermotoga
27	447.5	15.6	259	1 HIS6_ANASP	Q8yt31 anabaena sp
28	444.5	15.5	252	1 HIS6_STAAW	Q8nu13 staphylococ
29	441	15.3	198	1 HIS6_METH	Q27568 methanobact
30	432.5	15.0	251	1 HIS6_STRCO	Q98277 streptomyce
31	432.5	15.0	252	1 HIS6_BACSU	Q34727 bacillus su
32	432	15.0	208	1 HIS6_LISIN	Q92e86 listeria in
33	428	14.9	255	1 HIS6_NEIMA	Q9jvh5 neisseria m

RESULT 1

ID	HIS6_ARATH	STANDARD;	PRT;	592 AA.
AC	Q9S230; O80330;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Imidazole glycerol phosphate synthase hisH, chloroplast precursor			
DE	(IGP synthase) (IMP synthase) (IGPS) [Includes: Glutamine			
DE	amidotransferase (EC 2.4.2.-); Cyclase (EC 4.1.3.-)]			
GN	AT4G26900 OR F10M23.240.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RC	MEDLINE=98316711; PubMed=9654139;			
RA	Fujimori K., Ohta D.;			
RT	"An Arabidopsis cDNA encoding a bifunctional glutamine			
RT	amidotransferase/cyclase suppresses the histidine auxotrophy of a			
RT	Saccharomyces cerevisiae his7 mutant.";			
RN	FEBS Lett. 428:229-234(1998).			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RC	MEDLINE=20083488; PubMed=10617198;			
RA	Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,			
RA	Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,			
RA	Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,			
RA	Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,			
RA	Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,			
RA	Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,			
RA	Vos P., Honeisel J., Zimmermann W., Wedler H., Ridley P.,			
RA	Langham S.-A., McCullagh B., Bilham L., Robben J.,			
RA	Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,			
RA	Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,			
RA	Weitzensegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,			
RA	Holzner E., Brandt A., Peters S., van Staveren M., Dirkes W.,			
RA	Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,			
RA	Bernseis S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,			
RA	De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,			
RA	Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,			
RA	Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,			
RA	Pattett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,			
RA	Borkova D., Bloecher H., Scharfe M., Grimm M., Loehner T.-H.,			
RA	Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Partmann B., Ganderath K., Dauner D., Herzl A.,			
RA	Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,			
RA	Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,			
RA	Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,			
RA	Cheford F., Cooke R., Berger C., Monfort A., Casacuberta E.,			
RA	Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,			
RA	Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,			

ALIGNMENTS

34	427.5	14.9	251	1 HIS6_CHLTE	Q8kcb0 chlorobium
35	426	14.8	255	1 HIS6_NEIMB	Q9k0h4 neisseria m
36	426	14.8	258	1 HIS6_RHIME	Q92tb3 rhizobium m
37	423.5	14.7	252	1 HIS6_BACHD	Q9k626 bacillus ha
38	421	14.6	256	1 HIS6_RALSO	Q8x985 ralstonia s
39	416.5	14.5	261	1 HIS6_MYCLE	Q9x7c2 mycobacteri
40	410.5	14.3	261	1 HIS6_SYNY3	P74106 synchocyst
41	409	14.2	208	1 HIS6_LISMO	Q8y9g3 listeria mo
42	409	14.2	261	1 HIS6_AZOBH	P26721 azospirillu
43	407	14.1	214	1 HIS6_BACHD	Q9k624 bacillus ha
44	406	14.1	263	1 HIS6_RHILO	Q98ct1 rhizobium l
45	405	14.1	261	1 HIS6_BRUME	Q8ye37 bruceella me

Query Match 79.4%; Score 2284.5; DB 1; Length 592;
Best Local Similarity 80.7%; Pred.No. 4.7e-147;
Matches 435; Conservative 56; Mismatches 45; Indels 3; Gaps 2;

QY 21 AGAGDSVVTLLDYGAGNVRSVRNARTIGFDIKDVKPEDILNAKRLIFPGVGAFAPAM 80
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56 ASSTSDSVVTLLDYGAGNVRSRNALRHILGFISKDVTGTDLNADRLIFPGVGAFAPAM 115
QY 81 DVLIRKGLAEALCTYIQNDRPFPGICLGIQLLPESSEENGPIQGLGLIPGRVGRFPSSNG 140
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 DVLRRTGMAEALCKYLENDRPFGICLGIQLLPDSSEENGPAKGLGVIPQIVGRFASAG 175
QY 141 LRVPHICWHALDIKEGSAILDDVGNORHVFVHSYRA-NAEDNKEMISSTCSYGDDFIASI 199
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 IrvPHIGNALQVGKDEILDDVGNRHVFVHSYRAI PSDENKDWSISTCNYESFISSI 235
QY 200 QKNVNHAVQHPFKSGGVGLSILRRFLINADSFNKKQKPMNGKASKAKRVIACLDVran 259
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 RRGNNHAVQHPFKSGGEVGLSVLRRLFLHPKL--PATQKPMEGKASKAKRVIACLDVrTN 293
QY 260 DNGDLVVTTKGDQVDVERTTEENVNRNLGKPELAGOYLLDGADVSEFLNTITGRDPPLGD 319
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
294 DKGDLVVTTKGDQVDVREQNENEVRNLGKPVLDIAGQQYYKGDAEISFLNTITGRDPPLGD 353
QY 320 LPMLOVLQRASENVFPLTVGGGIRDTDANGRYYSLSLEVASFYFRSGADKVIGSDAVY 379
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 LPMIOVLRTSKNVFPLTVGGGIRDTDASGRYYSLSLEVAAYFRSGADKISIGSDAVS 413
QY 380 TABEYIKTVKTKSGKSIEIGISTVYIGNOAVVSIDPRRVLYRKPDVEFEFFKAIVSHPGPNG 439
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 AAEFIKSGVKTKSGKSLEQISRKYGNQAVVVSIDPRRVYNHPPDDVPYKIVRTNPGPNG 473
QY 440 EEFANTQCTVWGREGRPICAYELAKAVELGAGEILLNCIDCGOGKGDFIDI.LKLI.D 499
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474 EEFANTQCTVWGREGRPICAFELAKAVELGAGEILLNCIDCGOGKGDFIDILVKLI.D 533
QY 500 AVNIPIASSGADVADHFSEVFNETNASAAALAGIFHRKEVPIKAKVHELLKEGIEVRL 558
DB : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
534 SVCIPIASSGATGPDPHSEVFETKTWASAALAAGIFHRKEVPIQSVEHLQEERIEVRI 592

RESULT 2
HIS5 SCHPO STANDARD; PRT; 541 AA.
AC AC 034303; Q9Y7X3;
DT DT 16-OCT-2001 (Rel. 40, Created)
DT DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Imidazole glycerol phosphate synthase hisP (IGP synthase) (ImP
DE synthase) [IGPs] [Includes: Glutamine amidotransferase (EC 2.4.2.-);
DE Cyclase (EC 4.1.3.-)].
DN DN SPBC1A8.01C OR SPBC887.20C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RX STRAIN=972;
RN SEQUENCE FROM N.A.
RC MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltsellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.

CC -/- PATHWAY: Histidine biosynthesis; fifth step.
CC -/- INDUCTION: By amino acid starvation. It has a GCN4-dependent and
CC a GCN4-independent (basal) expression.
CC -/- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HISA / HISF
CC FAMILY.
CC -/- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69815; CAA9469.1; -.
DR EMBL; Z36117; CAA85211.1; -.
DR PIR; S46125; S46125.
DR PDB; 1JVN; 12-OCT-01.
DR SGD; S0000452; HIS7.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR006062; His_biosynth.
DR InterPro; IPR004651; HisF.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRFAMs; TIGR00735; hisF; 1.
DR PROSITE; PS00442; GATASE TYPE I; 1.
KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;
KW Lyase; Multifunctional enzyme; 3D-structure.
FT DOMAIN 1 210 AMIDOTRANSFERASE.
FT ACT_SITE 83 83 GATASE (BY SIMILARITY).
FT ACT_SITE 193 193 GATASE (BY SIMILARITY).
FT ACT_SITE 195 195 GATASE (BY SIMILARITY).
FT ACT_SITE 245 245 POTENTIAL.
FT ACT_SITE 404 404 POTENTIAL.
FT CONFLICT 54 54 G -> A (IN REF. 2).
SQ SEQUENCE 552 AA; 61068 MW; 046E11EA5F93ABA3 CRC64;

Query Match 47.4%; Score 1363; DB 1; Length 552;
Best Local Similarity 50.4%; Pred. No. 8.6e-85;
Matches 281; Conservative 91; Mismatches 149; Indels 36; Gaps 11;

Qy 28 VVTLLDYGAGNVRVNAIRTLGFDIKDVOKPED--ILNAKRLIFPGVGAFAPAMDVLR 85
Db 3 VVRVIDVESGNLSLTAIEHLGTEVQLVQKSPKDFNLSGTSRLILPQVNGYHFVNLFN 62
Qy 86 KGLAEALCTYIQNDPPLGLICGLQLLFESSENGPTQIGLIPGRVGRFESSNGLRVPH 145
Db 63 RGPEKPIREYESGKPTMGICVGLQALFAGSVESPKSTGLNYIDFKLSRPDDSE-KPVPE 121
Qy 146 IGHWALDIKGSAT-LDDVGNQHYVFVHSYRA--NAB-----DNKEWISSTCSYG-DDFI 196
Db 122 IGWNSCIPSNLFFGLDPY--KRYVFVHSPAALNSEKKLENLNDGKWKAKAKYSEEFI 179
Qy 197 ASTOKGNVHAVQHPKSGGVLSILRRFLNADSFNNKROK-----NWGKA 243
Db 180 AAVKNKNIFATQHPKSGKAGLVNFIENFL-----KQSPPIPNVSAEKKLLMNDYS 232
Qy 244 S-KLAKRVIACLDVRANDGDLVVTQGDQYDVRERTEENVRNLGKPVLAGQYLDGD 302
Db 233 NYGLTRRIIACLDVRTNDQGLVVTQGDQYDVRKSDGKVRNLGKPVLAQKYQGGAD 292
Qy 303 EVSFLNITGRDFPLGLDMLQVLQRASENVPVLTQGGIRPFTDANGRYYSLSLEAVE 362
Db 293 EVTFNLNITSFRDCPLKDTPLMLEVTKQAATVFVPLTVGGGKIDVDVGDGFKIPALEVASL 352
Qy 363 YFRSGADKVISGDAVVAETAEYIKTVK--TGKSSIEOISTVYGNQAVVVSIDPRRVYLRK 421
Db 353 YFRSGADKVISGDAVVAETAEYIKTVK--TGKSSIEOISTVYGNQAVVVSIDPRRVYLRK 412
Qy 422 PDEVEFKAIKVSHPGNPNGEYAWYQCTVNGGREGRPITGAYELAKAVEELGAGBILNCID 481

Db 413 QADTKNKVETBYPPGNGEYKWCYQCTIKGGRSDRLGVWELTRACEALGAGBILNCID 472
Qy 482 CDGQGGKFDIDILKILSDAVNIPVIASSGAGVADHSEVFNETNAAALAAAGIFHRKEVP 541
Db 473 KQGSNGSGYDLELIEHVKDAVKIPVIASSGAGVPEHFEAFELKTRADACLGAGNHRGEFT 532
Qy 542 IKAVKEHLLEKGEIVRL 558
Db 533 VNDVKVYLLEHGLKVRM 549

RESULT 4
HIS5 EMENI STANDARD; PRT; 553 AA.
ID HIS5 EMENI STANDARD; PRT; 553 AA.
AC Q9P4E9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase hisH (IGP synthase) (ImGP
DE synthase) (IGPs) [includes: Glutamine amidotransferase (EC 2.4.2.-);
DE Cyclase (EC 4.1.3.-)].
GN HISF.
OS *Emmericella nidulans* (*Aspergillus nidulans*).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; *Emmericella*.
RX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A234;
RX MEDLINE=21176419; PubMed=11277623;
RA Valerius O., Draht O., Kuebler E., Adler K., Hoffmann B., Braus G.H.;
RT "Regulation of hisH transcription of *Aspergillus nidulans* by adenine
RT and amino acid limitation";
RL Fungal Genet. Biol. 32:21-31(2001).
CC -/- FUNCTION: IGPs catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The glutamine amidotransferase domain
CC provides the ammonia necessary to the cyclase domain to produce
CC IGP and AICAR from PRFAR (By similarity).
CC -/- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulo-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -/- PATHWAY: Histidine biosynthesis; fifth step.
CC -/- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HISA / HISF
CC FAMILY.
CC -/- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF159463; AAF80376.1; -.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR006062; His_biosynth.
DR InterPro; IPR004651; HisF.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRFAMs; TIGR00735; hisF; 1.
DR PROSITE; PS00442; GATASE TYPE I; 1.
KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;
KW Lyase; Multifunctional enzyme.
FT DOMAIN 1 211 AMIDOTRANSFERASE.
FT ACT_SITE 81 81 GATASE (BY SIMILARITY).
FT ACT_SITE 194 194 GATASE (BY SIMILARITY).
FT ACT_SITE 196 196 GATASE (BY SIMILARITY).
FT ACT_SITE 241 241 POTENTIAL.
FT ACT_SITE 403 403 POTENTIAL.
SQ SEQUENCE 553 AA; 60359 MW; 8D48C524F3058FC2 CRC64;

Query Match 45.7%; Score 1315.5; DB 1; Length 553;
 Best Local Similarity 49.6%; Pred. No. 1.4e-81;
 Matches 276; Conservative 90; Mismatches 151; Indels 39; Gaps 11;

QY 29 VTLLDYGAGNVRVRAIRTLGDIKDQVQPEILNAKRLIFPGVGAFAFAMVDLTKGL 88
 DB 4 VHLDDYVAGNVRVRAIRTLGDIKDQVQPEILNAKRLIFPGVGAFAFAMVDLTKGL 88
 QY 89 AEALCTVIONDRPFLGICLQQLFSSSEBNGPIQGLLIPGRVGFSSNGLVRPHIGW 148
 DB 64 LQPIREHAGKPFMGICVGLQSLFSSSEBNGPIQGLLIPGRVGFSSNGLVRPHIGW 122
 QY 149 H-ALDIKEGSAIILDDVGNQ-----HYVFVHSYRANAE-----DNKEWISSTCSYGD 194
 DB 123 NSATDTR-----IDSTGQTFYGLSPSKYVYVHVAAPVEPGILEKDGMLVARRVYGE 177
 QY 195 -FIASTQKGNHVAQHPPEKSGGVLSILRFLNADS--FNNKRPQPMNGKAKLAKRVI 251
 DB 178 KFIGAIARDNIFATQHPPEKSGGGRPHPSRFLGRSSAPFCHIRDSILTGEKNGLTRRII 237
 QY 252 ACIDVRANDGDLVVTGDDQDVRETEEN---EVNRLGKPVLAGQYLLDGADEVSFLN 308
 DB 238 ACIDVRANDGDLVVTGDDQDVRETEEN---EVNRLGKPVLAGQYLLDGADEVSFLN 297
 QY 309 ITGFRDFPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLEVASSEYFRSGA 368
 DB 298 ITSFRNCLADPLMLLEIRLTSFVFPVPLTIGGGIRDFTDANGRYYSLEVASSEYFRSGA 357
 QY 369 DKVSGSDAVYTAEEYIKTG-VTKGKSSIRQISTVYGNQAVVSDIPRRVYLRKPDVEFP 427
 DB 358 DKVSGSDAVYTAEEYIKTG-VTKGKSSIRQISTVYGNQAVVSDIPRRVYLRKPDVEFP 417
 QY 428 KALKVSHPGNGBEYAWYQCTVNGREGRPDYGAVELAKAVEELGAGEILLNCIDCDGQK 487
 DB 418 RTIETFPNAGQNFQYQCTVNGREGRPDYGAVELAKAVEELGAGEILLNCIDCDGQK 477
 QY 488 GFDIDLKILSD-----AVNIPVIASSGAGVADHSEFSEVENETNAGAAAGIFHRKEVP 541
 DB 478 G-----SILTDQPPSKRAVKIPVIASSGAGMPHFEEVFDQTTDAAAGMFMHFRGEYT 531
 QY 542 IKAVKHEHLKEGIEVR 557
 DB 532 VGEVQKYLEDRGELVR 547

RESULT 5
 ID HIS6_ARCFU STANDARD; PRT; 271 AA.
 AC 029439;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Imidazole glycerol phosphate synthase subunit hsf (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hsf) (ImGP synthase subunit hsf) (IGPs subunit hsf).
 GN HISF OR AF0819

OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
 RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kervatage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 CC -I- FUNCTION: IGPs catalyzes the conversion of PRFAR and glutamine to
 CC IGP, AICAR and glutamate. The hsf subunit catalyzes the
 CC cyclization activity that produces IGP and AICAR from PRFAR using
 CC the ammonia provided by the hsf subunit (By similarity).
 CC -I- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
 CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
 CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
 CC aminimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
 CC -I- PATHWAY: Histidine biosynthesis; fifth step.
 CC -I- SUBUNIT: Heterodimer of hsf and hsf (By similarity).
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -I- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; A501047; AAB90415.1; -;
 DR PIR; C69352; C69352.
 DR TIGR; AF0819; -;
 DR HAMAP; MF_01013; -; 1.
 DR InterPro; IPR003009; FMN enzyme.
 DR InterPro; IPR006062; His biosynth.
 DR InterPro; IPR004651; HisF.
 DR Pfam; PF00977; His biosynth; 1.
 DR TIGRFAMs; TIGR00735; hsf; 1.
 DR Histidine biosynthesis; Lyase; Complete proteome.
 KW ACT SITE 11 11 POTENTIAL.
 FT ACT SITE 134 134 POTENTIAL.
 FT ACT SITE 134 134 POTENTIAL.
 SQ SEQUENCE 271 AA; 29861 MW; 91A5EA022AA3A5DA CRC64;

Query Match 18.4%; Score 529.5; DB 1; Length 271;
 Best Local Similarity 41.6%; Pred. No. 7.5e-29;
 Matches 131; Conservative 45; Mismatches 92; Indels 47; Gaps 8;

QY 246 LAKRVIACLDVRANDGDLVVTGDDQDVRETEENENLGPVELAGQYLLDGADEVS 305
 DB 2 LAKRIIPCLDVTI-DESEARVVGK-----VEFVNLRDAGDPVLAKEYDEEGDELV 52
 QY 306 FLNITGFRDFPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLEVASSEYFR 365
 DB 53 FLDTA---SPGRRTMIDVIERTAEQVFPFTVGGGKSIEDIN-----TILS 98
 QY 366 SGADKVSIGSDAVYTAEEYIKTGVTGKSSIRQISTVYGNQAVVSDIPRRVYLRKPD- 424
 DB 99 AGADKVSINTAAVKNPE-----FVREAADIFGQCITVAIDCRNFDLSKGEY 146
 QY 425 -VEFKAIVKSHPGNGEEYAWYQCTVNGREGRPDYGAVELAKAVEELGAGEILLNCIDCD 483
 DB 147 IVELE-----DGTK-AYEWYIYGRKPGVDVAVVWAKRVEELGAGEILLTSMNRD 196
 QY 484 GQKGFIDILKILSDAVNIPVIASSGAGVADHSEFSEVENETNAGAAAGIFHRKEVP 543
 DB 197 GTYDGFIDITRKISEEVNIPVIASSGAGTKEHFEGFVEGKADACLAASIFHYREIGR 256
 QY 544 AVKHEHLKEGIEVRL 558
 DB 257 EIKVLAERGQVRL 271

RESULT 6
 HIS6_METKA
 ID HIS6_METKA STANDARD; PRT; 272 AA.

Q8TYM8;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HSF OR MK0173.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; A010316; AAM01390.1; -
CC HAMAP; MF_01013; -; 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His_biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His_biosynth; 1.
CC TIGRfams; TIGR00735; hisF; 1.
KW Histidine biosynthesis; lyase; Complete proteome.
FT ACT_SITE 12 12 POTENTIAL.
FT ACT_SITE 131 131 POTENTIAL.
SQ SEQUENCE 272 AA; 29848 MW; CA7F9010E3A1A0CC CRC64;

Query Match 18.3%; Score 527; DB 1; Length 272;
Best Local Similarity 41.5%; Pred. No. 1.le-28;
Matches 131; Conservative 42; Mismatches 91; Indels 52; Gaps 8;

QY 246 LAKRVIACLDVRANDGDLVTKGDQYDVRTEENEVRNLGKPVLAGQVYLDGADEV 305
DB 3 LAKRIIFCLDVK-----DGRVVKGRF-----RGLRDAGDPAELAHYHGHGADEIV 49
QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFPLTVGGGIRGFTDANGRYYSLEVASFYR 365
DB 50 FLDISA---SPEGRLMDVVRVTAEKVFIPMTVGGGISDVED-----FR 91
QY 366 -----SGADKVSIGSDAYTAEEYIKTVGKTKSSIEQISTVYGNQAVVVIDPRVYLK 421
DB 92 RALTAGADKVSNTVAENPE-----LISEAADIFGQCQVVAIDAKREPL-K 138
QY 422 PDEVEFKAIKVSHPGNPEEYAWYQCTVNGGREGRPITGAVELAKAKAVEELGAGELLNCID 481

DB 139 PEHEHV-----ADHIFSNDDGEYWFVYVRGGRPEVDLDAITWAKRYVEELGAGILLTSID 194
QY 482 CDGQGGKFDIDIKLISDAVNIPVIASSGAGVADHSEVFNETNAAALAAAGFFHKEVP 541
DB 195 ADGTQEGYDIELTREVCAVSPVIASGGCGHPKHWVFKEADADALAASIFHYGKFT 254
QY 542 IKAVKEHLKKEGIEVR 557
DB 255 IEEVKEHLAERGVRVR 270

RESULT 7
HIS6 METTH STANDARD; PRT; 274 AA.
AC C27358;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HSF OR MTH1343.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7155-7155(1997).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; A0000897; AAB85821.1; ALT_INIT.
CC HAMAP; MF_01013; -; 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His_biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His_biosynth; 1.
CC TIGRfams; TIGR00735; hisF; 1.
KW Histidine biosynthesis; lyase; Complete proteome.
FT ACT_SITE 11 11 POTENTIAL.
FT ACT_SITE 134 134 POTENTIAL.
SQ SEQUENCE 274 AA; 30463 MW; B80082BE452AC53 CRC64;

```
Query Match      18.1%; Score 522; DB 1; Length 274;
Best Local Similarity 40.6%; Pred. No. 2.5e-28;
Matches 127; Conservative 44; Mismatches 102; Indels 40; Gaps 7;

QY 246 LAKRVITACLDVRANDGDLVVTGQDQYDVRTEENEVRNLKPVLAGQYYLDGADSVS 305
DB 2 LAKRIIIPCLDVK-----DLQVNG-----RVVKGEFKQIRYAGDVELATRYEDGADIV 52

QY 306 FLNITGFRDPLGLDMLQVLRASENVFPLTVGGGIRDTDANGRYYSLSLEVASSEYPR 365
DB 53 FLDTITASHE---RRETHVIAETENFVFCVGGGIRKPED---YFKML-----K 98

QY 366 SGADKVSIGSDAVYTAEEYIKTVGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 99 AGADKCSNTAAIKNPE-----LINEASDLVGSQACVVAIDAKRRYIENPRES 146

QY 426 EFKAIVKSHPGNGEBEYAWYQCTVNGGREGRPITGAYELAKAVEELGAGEILLNCIDCDGQ 485
DB 147 DERFIIIEVDG-----YCWYECISYGGREFTGIDAVKWMAMECDQDQAGEILLTSMRDGT 201

QY 486 KGKFDIDLKILSDAVNIPIASSGAGVADHSEVENETNASAAAGIFHRKEVPIKAV 545
DB 202 KMGTDIPLTMTGENDLPIVSIAGSGVGEPEHIYEAFDTGKADAAALASIFHNEYPPVAV 261

QY 546 KEHLKKEGIEVRL 558
DB 262 KEVLRSGVPIRL 274

RESULT 8
HIS6 THETA
ID HIS6 THETA STANDARD; PRT; 253 AA.
AC QSR85;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HISF OR TTE2133.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae;
OX NCBI_TaxID=119072;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=1197336;
RA Bao O., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribose-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)]imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC
```

```
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE013160; AAM25298.1; -.
DR HAMAP; MF_01013; -. 1.
DR InterPro; IPR003009; FMN enzyme.
DR InterPro; IPR006082; His_biosynth.
DR InterPro; IPR004651; HisF.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRFAMs; TIGR00735; hisF; 1.
KW Histidine biosynthesis; Lyase; Complete proteome.
FT ACT_SITE 11 11 POTENTIAL.
FT ACT_SITE 130 130 POTENTIAL.
SQ SEQUENCE 253 AA; 27516 MW; 4CAA77896FA57C2B CRC64;

Query Match      17.7%; Score 510.5; DB 1; Length 253;
Best Local Similarity 41.2%; Pred. No. 1.3e-27;
Matches 129; Conservative 42; Mismatches 81; Indels 61; Gaps 7;

QY 246 LAKRVITACLDVRANDGDLVVTGQDQYDVRTEENEVRNLKPVLAGQYYLDGADSVS 305
DB 2 LAKRIIIPCLDVK-----DGRVVKGINF-----VNLKADAGDPVEIABRYNELGADELV 48

QY 306 FLNITGFRDPLGLDMLQVLRASENVFPLTVGGGIRDTDANGRYYSLSLEVASSEYPR 365
DB 49 FLDTITASYE---KXKIMIDVVRKTSKVFIPITVGGGISDIDIR-----EVLK 94

QY 366 SGADKVSIGSDAVYTAEEYIKTVGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 95 AGADKVSINTQAV-----KQPTLIQALRFSGQCVVVAIDAKK----- 133

QY 426 EFKAIVKSHPGNGEBEYAWYQCTVNGGREGRPITGAYELAKAVEELGAGEILLNCIDCDGQ 485
DB 134 -----RPDGTGYNVY---INGRINTGLDAVEWAKVKVLDGAGEILLTSMRDGT 180

QY 486 KGKFDIDLKILSDAVNIPIASSGAGVADHSEVENETNASAAAGIFHRKEVPIKAV 545
DB 181 KQGYDIELTRLISEAVSIPIASGGAGKPEHREKFTQKADAAALASVPHYGELDIKEL 240

QY 546 KEHLKKEGIEVRL 558
DB 241 KRYLKDEGIPVRL 253

RESULT 9
HIS6 METJA
ID HIS6 METJA STANDARD; PRT; 272 AA.
AC Q57854;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HISF OR MJ0411.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock J.F., Weidman J.L., Nguyen D.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
```


QY 543 KAVKEHLKEGIEVR 557
 Db 236 PELKEYLLERGINVR 250

Matches 122; Conservative 48; Mismatches 100; Indels 45; Gaps 7;

QY 246 LAKRVIACLDVRANDNGDLVVTGKQYDVRTEENRNLGKPVLELAGQYLDGADSVS 305
 Db 2 LTKRIIPCLDVTLDVDRAGGCVKGVFVDLKE-----AGDPVELAKRYNEDGADLV 52

QY 306 FLNITGFRDFPLGDLPLQVLRASENVPLVTGGGIRHDTDFDANGRYYSSEVASEYPR 365
 Db 53 FLDTITASAH--GRTMIDVIERTADVEPIPLTVGGGI-----SSDAIRQLR 98

QY 366 SGADKVSIGSDAVYTAEEYIKTVKTKSSISQISTVYGNQAVVVISIDPR--VYLRKPD 423
 Db 99 AGADKVSNTSAVKNPD-----FIKSSDIFGAQCVITDAICRRNTDKRNP 146

QY 424 EYEFKRAIKVSHPGNPGEEYAVYQCTVNGREGRPICAYELAKAVELGAGEILLNCIDC 483
 Db 147 KT---VLELEDGTP-----AWYEVVYGGREATGIDAVQWAKKAELSGSEILLTSMRD 198

QY 484 GQKGFIDILIKLIDAVNIPVASSGAGVADHFSEVFNETHASAAALAGIFHRKEVPIK 543
 Db 199 GTCAGYDLPIITKLSELDPIITASGGVGNPQHIVEGSEGRADALAAASIFHFSEYSIW 258

QY 544 AVKEHLKEGIEVR 558
 Db 259 EVKEYLREREIPVRL 273

RESULT 12
 HIS6_METMA STANDARD; PRT; 273 AA.

AC Q8PW92;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase subunit hisF) (IGPS subunit hisF).
 GN HISF OR MMI704.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartesch T., Merkl R., Schmitz R.A., Martinez-Arias R., Henne A., Waeber A., Baumeister S., Jacob C., Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S., Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P., Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina mazei: evidence for lateral gene transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to IGP, AICAR and glutamate. The hisF subunit catalyzes the cyclization activity that produces IGP and AICAR from PRFAR using the ammonia provided by the hisH subunit (By similarity).
 CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-aminoimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
 CC -1- PATHWAY: Histidine biosynthesis; fifth step.
 CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE010714; AAM03985.1;
 CC HAMAP: MF 01013; 1.
 CC InterPro: IPR003009; FMN enzyme.
 CC InterPro: IPR006062; His_Biosynth.
 CC InterPro: IPR004651; HisF.
 CC Pfam: PF00977; His_Biosynth; 1.
 CC TIGRFAMs: TIGR00735; hisF; 1.
 CC Histidine biosynthesis; Lyase; Complete proteome.
 CC ACT_SITE 11 11 POTENTIAL.
 CC ACT_SITE 134 134 POTENTIAL.
 CC SEQUENCE 273 AA; 29767 MW; B3BAED4D8602C06A CRC64;

Query Match 17.2%; Score 493.5; DB 1; Length 273;
 Best Local Similarity 38.7%; Pred. No. 2.1e-26;

QY 543 KAVKEHLKEGIEVR 557
 Db 236 PELKEYLLERGINVR 250

Matches 122; Conservative 48; Mismatches 100; Indels 45; Gaps 7;

QY 246 LAKRVIACLDVRANDNGDLVVTGKQYDVRTEENRNLGKPVLELAGQYLDGADSVS 305
 Db 2 LTKRIIPCLDVTLDVDRAGGCVKGVFVDLKE-----AGDPVELAKRYNEDGADLV 52

QY 306 FLNITGFRDFPLGDLPLQVLRASENVPLVTGGGIRHDTDFDANGRYYSSEVASEYPR 365
 Db 53 FLDTITASAH--GRTMIDVIERTADVEPIPLTVGGGI-----SSDAIRQLR 98

QY 366 SGADKVSIGSDAVYTAEEYIKTVKTKSSISQISTVYGNQAVVVISIDPR--VYLRKPD 423
 Db 99 AGADKVSNTSAVKNPD-----FIKSSDIFGAQCVITDAICRRNTDKRNP 146

QY 424 EYEFKRAIKVSHPGNPGEEYAVYQCTVNGREGRPICAYELAKAVELGAGEILLNCIDC 483
 Db 147 KT---VLELEDGTP-----AWYEVVYGGREATGIDAVQWAKKAELSGSEILLTSMRD 198

QY 484 GQKGFIDILIKLIDAVNIPVASSGAGVADHFSEVFNETHASAAALAGIFHRKEVPIK 543
 Db 199 GTCAGYDLPIITKLSELDPIITASGGVGNPQHIVEGSEGRADALAAASIFHFSEYSIW 258

QY 544 AVKEHLKEGIEVR 558
 Db 259 EVKEYLREREIPVRL 273

RESULT 12
 HIS6_METMA STANDARD; PRT; 273 AA.

AC Q8PW92;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase subunit hisF) (IGPS subunit hisF).
 GN HISF OR MMI704.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartesch T., Merkl R., Schmitz R.A., Martinez-Arias R., Henne A., Waeber A., Baumeister S., Jacob C., Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S., Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P., Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina mazei: evidence for lateral gene transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to IGP, AICAR and glutamate. The hisF subunit catalyzes the cyclization activity that produces IGP and AICAR from PRFAR using the ammonia provided by the hisH subunit (By similarity).
 CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-aminoimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
 CC -1- PATHWAY: Histidine biosynthesis; fifth step.
 CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE010714; AAM03985.1;
 CC HAMAP: MF 01013; 1.
 CC InterPro: IPR003009; FMN enzyme.
 CC InterPro: IPR006062; His_Biosynth.
 CC InterPro: IPR004651; HisF.
 CC Pfam: PF00977; His_Biosynth; 1.
 CC TIGRFAMs: TIGR00735; hisF; 1.
 CC Histidine biosynthesis; Lyase; Complete proteome.
 CC ACT_SITE 11 11 POTENTIAL.
 CC ACT_SITE 134 134 POTENTIAL.
 CC SEQUENCE 273 AA; 29767 MW; B3BAED4D8602C06A CRC64;

Query Match 17.2%; Score 493.5; DB 1; Length 273;
 Best Local Similarity 38.7%; Pred. No. 2.1e-26;

```
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; AE013407; AA031400.1; -.
CC DR HAMAP; MF 01013; -. 1.
CC DR InterPro; IPR003009; FMN enzyme.
CC DR InterPro; IPR006062; His biosynth.
CC DR InterPro; IPR004651; HisF.
CC DR Pfam; PF00977; His biosynth; 1.
CC DR TIGRFAMs; TIGR00735; hisF; 1.
CC KW Histidine biosynthesis; Lyase; Complete proteome.
CC FT ACT SITE 11 11 POTENTIAL.
CC FT ACT SITE 134 134 POTENTIAL.
CC SQ SEQUENCE 273 AA; 29690 MW; FCC6AD1B6377918 CRC64;

Query Match 17.0%; Score 489.5; DB 1; Length 273;
Best Local Similarity 38.7%; Pred. No. 3.9e-26;
Matches 121; Conservative 48; Mismatches 103; Indels 41; Gaps 6;

QY 246 LAKRVIACLDVRANDGDLVVTGDDQYDVRTEENEVRNLGKPVLAGQYLDGADSVS 305
DB 2 LTRKRIIPCLDVTLDAGGCVKGVFVDLKE-----AGDPVELAKRYNEDGDELV 52
QY 306 FLNITGPRDFPLGDLPLQVLRASENVFVPLTVGGGIRDTDANGRYYSLSLEVASVEYFR 365
DB 53 FLDITASAH---GRETMIDVIERTADEVFIPITVGGGI-----SSIEAIRQILR 98
QY 366 SGADKVSIGSDAVTAEYIKTGVTGKSSIEQISTVYGNQAVVSDIPRRVLRKPDEV 425
DB 99 AGADKVSINTSAVKNP-----FIKSSDIFGAQCIIVTADICKR-NTNVKNDP 145
QY 426 EFRAIKVSHPGNGBEYAWYQCTVNGGREGRPITGAYELAKAVELGAGEILLNCIDCDGQ 485
DB 146 DKTILELEDGTP-----AWTEVYIYGRKNTGIDAVQWAKRAELGSGEILLTSMRDGT 200
QY 486 GKGFDDILKILSDAVNIPVIASSGAGVADHFESEVFNETNASAAAGIFHRKEVPIKAV 545
DB 201 CAGYDIPITRKLSELDIPAIASGGVGNPQHIYEGFSDGKADAAALAAASIFHGEYSIQEV 260
QY 546 KEHLKKEGIEVRL 558
DB 261 KEFLKPKIPVRL 273

RESULT 13
HIS6_CLOAB STANDARD; PRT; 253 AA.
AC Q97KH8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase subunit hisF) (IGPS subunit hisF).
DE HISF OR CAC0941.
GN Clostridium acetobutylicum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."
RT J. Bacteriol. 183:4823-4838 (2001).
CC -1- FUNCTION: IGP synthase catalyzes the conversion of PRFAR and glutamine to IGP, AICAR and glutamate. The hisF subunit catalyzes the cyclization activity that produces IGP and AICAR from PRFAR using the ammonia provided by the hisH subunit (By similarity).
```

```
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-ylamino)methylidenamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AE007609; AAK78917.1; -.
CC DR PIR; B97016; B97016.
CC DR HAMAP; MF 01013; -. 1.
CC DR InterPro; IPR003009; FMN enzyme.
CC DR InterPro; IPR006062; His biosynth.
CC DR InterPro; IPR004651; HisF.
CC DR Pfam; PF00977; His biosynth; 1.
CC DR TIGRFAMs; TIGR00735; hisF; 1.
CC KW Histidine biosynthesis; Lyase; Complete proteome.
CC FT ACT SITE 11 11 POTENTIAL.
CC FT ACT SITE 130 130 POTENTIAL.
CC SQ SEQUENCE 253 AA; 27541 MW; 9B52D17AACF14FE0 CRC64;

Query Match 16.7%; Score 481.5; DB 1; Length 253;
Best Local Similarity 37.4%; Pred. No. 1.2e-25;
Matches 117; Conservative 44; Mismatches 91; Indels 61; Gaps 7;

QY 246 LAKRVIACLDVRANDGDLVVTGDDQYDVRTEENEVRNLGKPVLAGQYLDGADSVS 305
DB 2 LTRKRIIPCLDVTLDAGGCVKGVFVDLKE-----VNLKDVGPVIEADFYNREGADEIV 48
QY 306 FLNITGPRDFPLGDLPLQVLRASENVFVPLTVGGGIRDTDANGRYYSLSLEVASVEYFR 365
DB 49 FLDISATNE---GRNTMIDVVRRTAEKVFIPITVGGGIRDVD-----FKNILR 94
QY 366 SGADKVSIGSDAVTAEYIKTGVTGKSSIEQISTVYGNQAVVSDIPRRVLRKPDEV 425
DB 95 AGADKISINSALRNPE-----LINEAKKFGSCQVVAVDARR-----133
QY 426 EFRAIKVSHPGNGBEYAWYQCTVNGGREGRPITGAYELAKAVELGAGEILLNCIDCDGQ 485
DB 134 -----NIETGGFNVPINGRVDTSIDAIETWVKCESLGAIGEILLTSMADGDT 180
QY 486 GKGFDDILKILSDAVNIPVIASSGAGVADHFESEVFNETNASAAAGIFHRKEVPIKAV 545
DB 181 KNGYDVELTGAVCDVAVRIPVIASGGCGRLDEDFEIKKTSADAALVASLFHYKEYTVGEV 240
QY 546 KEHLKKEGIEVRL 558
DB 241 KNYLRENSIQVRL 253

RESULT 14
HIS6_CORGL STANDARD; PRT; 258 AA.
ID HIS6_CORGL
AC Q31139;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase subunit hisF) (IGPS subunit hisF).
DE HISF OR CGL2094.
GN Corynebacterium glutamicum (Brevibacterium flavum).
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:23:47 ; Search time 40 seconds
(without alignments)

3599.834 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877

Sequence: 1 HEKELASTKPNQGFRIKAL.....EVIKAVKSHLLKEGIEVRL 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2270.5	78.9	592	10 Q8GX14	Q8GX14 arabidopsis
2	1192.5	41.4	627	3 Q9HFV7	Q9HFV7 candida alb
3	484.5	16.8	258	16 Q8FNZ9	Q8FNZ9 corynebacte
4	476	16.5	261	2 Q8GDQ6	Q8GDQ6 heliobacill
5	439	15.3	256	16 Q8GF7	Q8GF7 bifidobacte
6	438	15.2	203	16 Q8R883	Q8R883 thermoaer
7	432.5	15.0	254	16 Q8ESS2	Q8ESS2 oceanobacil
8	424.5	14.8	252	16 Q8CQ92	Q8CQ92 staphylococ
9	421	14.6	251	16 Q8DTR3	Q8DTR3 streptococ
10	412.5	14.3	281	16 Q8DJN7	Q8DJN7 synecococc
11	407	14.1	201	16 Q8KF56	Q8KF56 chlorobium
12	405	14.1	261	16 Q8FY07	Q8FY07 brucella su
13	396	13.8	201	16 Q8ESS0	Q8ESS0 oceanobacil
14	394	13.7	201	2 Q939J6	Q939J6 campylobact
15	393	13.7	202	17 Q8TS91	Q8TS91 methanosarc
16	390.5	13.6	209	16 Q8DIP5	Q8DIP5 synecococc

17	381	13.2	207	17 Q8PVD5	Q8PVD5 methanosarc
18	378.5	13.2	257	16 Q8D8Q5	Q8D8Q5 vibrio vuln
19	378	13.1	201	16 Q8DTR1	Q8DTR1 streptococ
20	358.5	12.5	258	16 Q8FG48	Q8FG48 escherichia
21	357.5	12.4	257	16 Q8CX42	Q8CX42 shewanella
22	354	12.3	192	16 Q8CTV0	Q8CTV0 staphylococ
23	344	12.0	200	16 Q8PLG8	Q8PLG8 xanthomonas
24	343.5	11.9	204	16 Q8D8Q3	Q8D8Q3 vibrio vuln
25	339	11.8	200	16 Q8P9P1	Q8P9P1 xanthomonas
26	336.5	11.7	215	16 Q8G4S6	Q8G4S6 bifidobacte
27	335.5	11.7	221	16 Q8EFB4	Q8EFB4 shewanella
28	328	11.4	216	16 Q8FY09	Q8FY09 brucella su
29	319	11.1	211	16 Q8FNZ6	Q8FNZ6 corynebacte
30	317.5	11.0	204	2 Q8KIU3	Q8KIU3 pseudomonas
31	313.5	10.9	200	17 Q8TV83	Q8TV83 methanopyru
32	265	9.2	251	2 Q8KIS2	Q8KIS2 pseudomonas
33	216	7.5	104	2 Q8RPQ6	Q8RPQ6 thermoaer
34	213	7.4	166	2 Q8VNA3	Q8VNA3 rhizobium e
35	208.5	7.2	260	2 Q8KIU2	Q8KIU2 pseudomonas
36	170.5	5.9	240	16 Q8ESS1	Q8ESS1 oceanobacil
37	151	5.2	114	2 Q9F8I8	Q9F8I8 carboxydoth
38	148	5.1	502	16 Q8RJ20	Q8RJ20 streptomyc
39	145	5.0	479	2 Q939K5	Q939K5 propionibac
40	139	4.8	816	10 Q9LJG4	Q9LJG4 arabidopsis
41	138	4.8	808	10 Q94A27	Q94A27 arabidopsis
42	130.5	4.5	246	16 Q8FG49	Q8FG49 escherichia
43	129.5	4.5	234	16 Q8CQ93	Q8CQ93 staphylococ
44	128	4.4	239	16 Q8DTR2	Q8DTR2 streptococ
45	127.5	4.4	937	17 Q8PTT2	Q8PTT2 methanosarc

ALIGNMENTS

RESULT 1

Q8GX14 ID Q8GX14 PRELIMINARY; PRT; 592 AA.

AC Q8GX14; DT 01-MAR-2003 (T-REMBLrel. 23, Created)

DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-REMBLrel. 23, Last annotation update)

DE Putative glutamine amidotransferase/cyclase.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosid II; Brassicales; Brassicaceae; Arabidopses.

OC NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,

RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,

RA Hayashizaki Y., Shinozaki K.;

RT "Arabidopsis thaliana full-length cDNA."

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK118498; BAC43102.1; -

KW Transferase; Glutamine amidotransferase.

SQ SEQUENCE 592 AA; 64162 MW; FDD7F2B2F70CD515 CRC64;

Query Match 78.9%; Score 2270.5; DB 10; Length 592;
Best Local Similarity 80.5%; Pred. No. 1.7e-146;
Matches 434; Conservative 56; Mismatches 46; Indels 3; Gaps 2;

Qy	21	AGAGGSVVTLLDYGAGNVSVRNATRTGFDIKDVQKPEDILNAKRLIPPGVGAFAPAM 80	
Db	56	ASSTSDSVVTLLDYGAGNVSIRNARHLGFSIKDVQTPGDILNADRLIPPGVGAFAPAM 115	
Qy	81	DVLRKGLAEALCTYIQNDRPPFLGICLGLQLLPESSEENGPIQGLGIPGRVGRFESSNG 140	
Db	116	DVLRRTGMAEALCKYIENRPPFLGICLGLQLLPDSSEENGPIQGLGIPGRVGRFESSNG 175	
Qy	141	LRYPHIGWALDIKEGSAILDDVGNQHVYFVHSYRA-NAEDNKWEISSTCSYGDFTASI 199	

Db 176 IRVPHIGWALQVKDSEILDDVGNRRHVFVHSYRAIPSDENKDWISSTCNYGESFISII 235
 Qy 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKVIACLDVRAN 259
 Db 236 RRGNVHAVQHPKSGVGLSVLRRLHPKL--PATQKPMNGKASKLAKVIACLDVRN 293
 Qy 260 DNGDLVVTGQDQDVDRTERTEENVRNLGKVELAGQYLLDGADEVFLNITGFRDPLGD 319
 Db 294 DKGDLVVTGQDQDVDRTERTEENVRNLGKVELAGQYLLDGADEVFLNITGFRDPLGD 353
 Qy 320 LPMQLVLRASENVFPLTVGGGIRDPDANGRYYSLEVASVYFRSGADKVSIGSDAVY 379
 Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDPDANGRYYSLEVASVYFRSGADKVSIGSDAVS 413
 Qy 380 TAEEYIKTVKTKSSIEQISTVYGNQAVVSDIPRRVYLRKDEVEFKAIKVSHPGNG 439
 Db 414 AAEEFKSGVTKGSSLEQISRVYGNQAVVSDIPRRVYLRKDEVEFKAIKVSHPGNG 473
 Qy 440 EEVAVYQCTVNGGREGRPICAYELAKAVEELGAGEILLNCIDCDGQGGKGFIDILKILSD 499
 Db 474 EEVAVYQCTVNGGREGRPICAYELAKAVEELGAGEILLNCIDCDGQGGKGFIDILKILSD 533
 Qy 500 AVNIPVIASSGAGVADHFSVFNETNASALAAGIFHRKEVPIKAVKEHLKSGIEVRL 558
 Db 534 SVGIPVIASSGAGTPDHFSEVFETNASALAAGIFHRKEVPIQSVKEHLQEEIEVRI 592

RESULT 2
 Q9HFV7 ID Q9HFV7 PRELIMINARY; PRT; 627 AA.
 AC Q9HFV7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Imidazole glycerol phosphate synthase.
 GN HIS7.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Day T.W., Davissou V.J.;
 RT "Cloning and Characterization of the CaHIS6 and CaHIS7 Genes from the
 Fungal Pathogen Candida albicans.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF290177; ANGI7872.1; -;
 DR HSP; Q9X0C6; 1THF.
 DR InterPro; IPR000991; GATase_1.
 DR InterPro; IPR004651; HisP.
 DR InterPro; IPR006062; His_biosynth.
 DR Pfam; PF00117; GATase; 1.
 DR Pfam; PF00977; His_biosynth; 1.
 DR TIGRFAMs; TIGR00735; hisP; 1.
 DR PROSITE; PS00442; GATASE TYPE I; 1.
 SQ SEQUENCE 627 AA; 69605 MW; D5A51D5D3A6A34C CRC64;

Query Match 41.4%; Score 1192.5; DB 3; Length 627;
 Best Local Similarity 43.6%; Pred. No. 7e-73;
 Matches 272; Conservative 104; Mismatches 147; Indels 101; Gaps 19;
 Qy 29 VTLDDYAGNVRSVRNRI--GFDIKDQVQKPEDILN---AKRLFFPGVGAFAFAMDV 82
 Db 5 IHIIDVSGNLQSLNKAIRDSNYLIKFTNEDQFDINDSQIEKLIFPGVGNFHFVKQ 64
 Qy 83 LIRKGLAEALCTYQNDPRFLGICGLQLLFESSENGPIQGLGLIPG-----130
 Db 65 LNERKLNHLKSYIKQDPLMGICVGIQSIFHESESPNKGKGLGLLDNDNDDEKKL 124
 Qy 131 RVGRFESSN-----GLR--VPHIGWAL--DIKESGAILDDVGNQ-----HV 168
 Db 125 KLYKFDNDDEKFKIRGKSVPHIGWNNIHI-----IIN--GKQTTSLYGLNKIDKY 177

Qy 169 YFVHSTRANAEDN-----KEWISSTCSYGD--FIASIOKGNHVAOPHPE 212
 Db 178 YFVHVAALIKNNNDIEIKKFKIDLSKGDWLAISQYSGSEKFIKSAIKNNLPATQFHP 237
 Qy 213 KSGGVLGSLIRPLN-----ADSFNNKRQKPMNGKASKLAKVIACL 254
 Db 238 KSGIVGLKIKNFKLNGEKNPPTPTPTNNLAQGVSDSSSTLDVETTLTGLTRRIACL 297
 Qy 255 DYRNDNGDLVVTGQDQVDRER-----TEENVRNLGKVELAGQYLLDGADE 303
 Db 298 DVRTNDGDLVVTGQDQVDRER-----TEENVRNLGKVELATKYNQGADE 357
 Qy 304 VSFNITGFRDPLGDLPMQLVLRASENVFPLTVGGGIRDPD--ANGRYYSLEVAS 362
 Db 358 IIFLNITSFRNPLKDLPMQLVLRASENVFPLTVGGGIRDPD--ANGRYYSLEVAS 417
 Qy 363 YFRSGADKVSIGSDAVYTAEE--YIKTVKTKSSIEQISTVYGNQAVVSDIPRRVYLRK 421
 Db 418 YFQSGADKVSIGSDAVYTAEE--YIKTVKTKSSIEQISTVYGNQAVVSDIPRRVYLRK 477
 Qy 422 P--DEVEFKAIKVSHP---GNGEYAWYQCTVNGGREGRPICAYELAKAVEELGAGEIL 476
 Db 478 PTTETTTMTQIKITDPSQYGPNGEQYQYQVTSQGGKRIHELGAELCLACEKLGAGEIL 537
 Qy 477 LNCIDCDGQGGKGFIDILKILSDAVNIPVIASSGAGVADHFSVFN--ETNASALAAGIF 535
 Db 538 LNSIDHSGNKGFLNQLLOQIKSVIPVIASSGAGNPHQFQGVFEMDCGIDAALGAGLF 597
 Qy 536 HRKEVPIKAVKEHLKSG--IEVRL 558
 Db 598 HRGEYTVNQVKYLOQEAQKMDVRL 621

RESULT 3
 Q8FNZ9 ID Q8FNZ9 PRELIMINARY; PRT; 258 AA.
 AC Q8FNZ9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cyclase HisP.
 GN HISF OR CE1994.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=152794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
 RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.;
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005220; BAC18804.1; -;
 KW Complete proteome.
 SQ SEQUENCE 258 AA; 27106 MW; 7E57F451BFBDF8A CRC64;

Query Match 16.8%; Score 484.5; DB 16; Length 258;
 Best Local Similarity 38.8%; Pred. No. 3.7e-25;
 Matches 121; Conservative 51; Mismatches 83; Indels 57; Gaps 9;
 Qy 246 LAKRVITACLDVRANDNGDLVVTGQDQVDRTERTEENVRNLGKVELAGQYLLDGADEV 305
 Db 3 VAIRVIPCLDV---DNGRVV--KGVNFE-----NLRDAGDPLAKRYGEGADELT 49
 Qy 306 FLNITGFRDPLGDLPMQLVLRASENVFPLTVGGGIRDPD--ANGRYYSLEVASVYFR 365
 Db 50 FLDVSASKD---GRGTMLDVRRTAQIFIPITVGGV-----SVEDVQLLR 95
 Qy 366 SGADKVSIGSDAVYTAEEYIKTVKTKSSIEQISTVYGNQAVVSDIPRRVYLRKDE 425
 Db 96 AGADKVSNTSATARPE-----LLSELSQRFQAQCVLSVDARRVPAGEA---140

Qy 426 EFKAIVSHFGPNGEYVAMVQCTVNGRGPRIGAYELAKAVBELGAGEILLNCIDCDGQ 485
Db 141 -----PQSG-----FEVTHGCTRSAGLDVAWEAITGKGVGEILLNSMDGDT 186
Qy 486 GKGFIDILKILSDAVNIPVIASSGAGVADHPSFVFNETNASALAAGIFHRKEVPKAV 545
Db 187 KNGFDLELLKRAAVSIPVIASSGAGTAEHFPFPAV-RAGANAVLAATIFHFGEVITTEV 245
Qy 546 KEHLKEGIEVR 557
Db 246 KDAIEKAGFEVR 257

RESULT 4

Q8GDQ6 PRELIMINARY; PRT; 261 AA.
AC Q8GDQ6; 2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hlep (Fragment)
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22337798; PubMed=12446909;
RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
RA Blankenship R.E.;
RT "Whole-genome analysis of photosynthetic prokaryotes.";
RL Science 298:1616-1620(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
RA Gerdes S., Kyrpides N., Overbeek R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AV142920; AA87524.1; --
FT NON TER 261
SQ SEQUENCE 261 AA; 28552 MW; 69B03C9560582690 CRC64;

Query Match 16.5%; Score 476; DB 2; Length 261;

Best Local Similarity 38.8%; Pred. No. 1.4e-24;

Matches 121; Conservative 41; Mismatches 94; Indels 56; Gaps 7;

Qy 246 LAKRVIACLDVRANDNGDLVVTGKQDYDVRTEENEVRNLGKPVELAGQYLDGAEV 305
Db 2 LAKRIIPCLDVHGR-----VWKG-----TNFVNLRDAGDPVELAAAYDKGADLV 48
Qy 306 FLNITGFRDPLGLPMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLEVASFYR 365
Db 49 FLDITASSD---GRAIMLDVRRTAEBVFIPFTVGGGLR-----TVEDIREMLK 94

Qy 366 SGADKVSIGSDAVYTAEEYIKTVGKTSIEQISTVYGNQAVVVSIDPRRVYLKRPDE 425
Db 95 AGADKISLNTSAVQTPQ-----LIGDSAWKFGSQCVVAIDARR---RRDEE- 138

Qy 426 EFKAIVSHFGPNGEYVAMVQCTVNGRGPRIGAYELAKAVBELGAGEILLNCIDCDGQ 485
Db 139 -----GRLEGHEVYTHGGRKPTGIDVLEWARKVBEKGCEILLTSMDDGDT 185

Qy 486 GKGFIDILKILSDAVNIPVIASSGAGVADHPSFVFNETNASALAAGIFHRKEVPKAV 545
Db 186 KDGFDIPLTRAVSAVKIPVIASSGGVGNLEHICEGLTAGKADALAASIFHYKEYTIRET 245

Qy 546 KEHLKEGIEVR 557

Db 246 KEYLRGKGHVHR 257

RESULT 5

Q8G6F7

ID Q8G6F7 PRELIMINARY; PRT; 256 AA.
AC Q8G6F7;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hlep cyclase.
GN HISP OR BL0885.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Beiger B.,
RA Pessi G., Zwhalen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL: AE014691; AA24505.1; --
KW Complete proteome.
SQ SEQUENCE 256 AA; 27551 MW; 7CE149FCB0759B5C CRC64;

Query Match 15.3%; Score 439; DB 16; Length 256;

Best Local Similarity 35.0%; Pred. No. 4.6e-22;

Matches 111; Conservative 46; Mismatches 92; Indels 68; Gaps 9;

Qy 246 LAKRVIACLDVRANDNGDLVVTGKQDYDVRTEENEVRNLGKPVELAGQYLDGAEV 305
Db 3 LAVRVIPCLDVRAGR-----VWKGVHF-----NLKDAGDPVELAAEYVQRADET 49

Qy 306 FLNIT---GFRDPLGLPMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLEVAS 362
Db 50 FLDVTASSSHRN-----TMDVVSRTABQVPIPTVGGVTRPEDVD-----S 92

Qy 363 YFRSGADKVSIGSDAVYTAEEYIKTVGKTSIEQISTVYGNQAVVVSIDPRRVYLKRP 422
Db 93 LLRCGADKGVNTAAI-----NDPSLISRVADRFGQVQLVLSVDARR----- 134

Qy 423 DEVEFKAIKSHPGPNGEY---AWYQCTVNGRGPRIGAYELAKAVBELGAGEILLNCI 480
Db 135 -----EKGEQTSQGFVETMGGRKSTGIDAIWVVKRAEQLGAGEILLNSM 180

Qy 481 DCDGQCKGFDILKILSDAVNIPVIASSGAGVADHPSFVFNETNASALAAGIFHRKEV 540
Db 181 DAGTGKGFDELMIRAVRKEVKIPIIASGAGKVEDFPFPAI-BAGADAVLAASVFFHYGIL 239

Qy 541 PIKAVKEHLKEGIEVR 557

Db 240 TIADVKAELKKGHTYR 256

RESULT 6

Q8R883

ID Q8R883 PRELIMINARY; PRT; 203 AA.

AC Q8R883;

DT 01-JUN-2002 (T-EMBLrel. 21, Created)

DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)

DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)

DE Glutamine amidotransferase.

GN HISH OR TTE2135.

OS Thermoanaerobacter tengcongensis.

OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;

OC Thermoanaerobacteriaceae; Thermoanaerobacter.

OX NCBI_TaxID=119072;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MB4 / JCM 11007;

RX MEDLINE=21992816; PubMed=11997336;

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,

RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UA159 / ATCC 700610 / Serotype C;
 RX MEDLINE=22295063; PubMed=12397186;
 RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.,
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 RT pathogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
 DR ENBL; AE014962; AAN58946.1; --
 KW Complete proteome.
 SQ SEQUENCE 251 AA; 26874 MW; 1C27D132D16BFBEA CRC64;

Query Match 14.6%; Score 421; DB 16; Length 251;
 Best Local Similarity 34.8%; Pred. No. 7.5e-21;
 Matches 110; Conservative 47; Mismatches 89; Indels 70; Gaps 10;

QY 246 LAKRVACLDVRANDGLVVTGDDQYDVRTEENRNLGKPVLAGOYLDGAEVS 305
 Db 2 LKRIIPCLDVK-----DGRVKGVF-----VNLTDVGPVDAARAYYAGCDELV 48
 QY 306 FLNITGPRDFPLGLDPLQVLRASENVFPLTVGGGIRDFTDANGRIYSSLEVSEYFR 365
 Db 49 FLDTATSD---NREITVDMVRHVADQVFPTVGGGIRSVDDMN-----KMLK 94
 QY 366 SGADKSVISGDAVYTAEEYIKTGKTSIEIQISTVYGNQAVVSDPRRVYLRKPDEV 425
 Db 95 AGADKVAVNSSAI-----ANPLIKDCAEKFGSCVQVVAIDAR---KEADD- 137
 QY 426 EPKAIKVSHPGNGEYVAVYQCTVNGGREGRPIGAYELAKAVELGAGEILLNCIDCGQ 485
 Db 138 -----SWH-VYVAGGRKDTGIDLAAVKEAVOLGAGEILLTSMOKDGT 179
 QY 486 GKGFIDIDLKILISDAVNPIVSIASSGAGVADHFSEVFNENASAAAGIFHRKEVPI--- 542
 Db 180 KSGFDLMLNAVLAQADIPIIASSGAGNMEHVEIFKTPATGALAAASIFHYGEVSADT 239
 QY 543 -KAVKEHLKKEGIEVR 557
 Db 240 KKAMKEH-----GIEVR 251

RESULT 10
 Q8DJN7
 ID Q8DJN7 PRELIMINARY; PRT; 281 AA.
 AC Q8DJN7
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE Inidazoleglycerol-phosphate synthase, cyclase subunit.
 GN HSF OR TLR1185.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1.";
 RL DNA Res. 9:123-130 (2002).
 DR ENBL; AP005373; BAC08737.1; --
 KW Complete proteome.
 SQ SEQUENCE 281 AA; 30011 MW; A8AB0E9F7430372 CRC64;

Query Match 14.3%; Score 412.5; DB 16; Length 281;
 Best Local Similarity 34.3%; Pred. No. 3.4e-20;
 Matches 108; Conservative 47; Mismatches 97; Indels 63; Gaps 7;

QY 246 LAKRVACLDVRANDGLVVTGDDQYDVRTEENRNLGKPVLAGOYLDGAEVS 305
 Db 3 LAKRIIPCLDVKAGR-----VKGGVNF-----VNLRDAGDPVELAQYNAAGADELV 49
 QY 306 FLNITGPRDFPLGLDPLQVLRASENVFPLTVGGGIRDFTDANGRIYSSLEVSEYFR 362
 Db 50 FLDTATHEARNI-----LIDVVTADQVFIPLTVGGGIQ-----SLTMKD 92
 QY 363 YFRSGADKSVISGDAVYTAEEYIKTGKTSIEIQISTVYGNQAVVSDPRRVYLRKP 422
 Db 93 LLRAGADKVSLSAAVRQPD-----LVNQASDRFGAQCIIVVAIDARREPDCA 140
 QY 423 DEVEFKAIKVSHPGNGEYVAVYQCTVNGGREGRPIGAYELAKAVELGAGEILLNCIDC 482
 Db 141 DQPR-----WQVYVGRGREATGLDAVAVEMAKRGAAGELLVTSMDA 182
 QY 483 DGOGKGFIDIDLKILISDAVNPIVSIASSGAGVADHFSEVFNENASAAAGIFHRKEVPI 542
 Db 183 DGTQAGYDLELTAERVEIPVSIASSGAGTCEHRAALVEGKAEALLASLHYGOLTI 242
 QY 543 KAVKEHLKKEGIEVR 557
 Db 243 AQIKGYLHQHQPVR 257

RESULT 11
 Q8KF56
 ID Q8KF56 PRELIMINARY; PRT; 201 AA.
 AC Q8KF56
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE Amidotransferase Hish.
 GN HSH OR CT0476.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OX Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
 DR ENBL; AS012824; AAM71718.1; --
 DR TIGR; CT0476; --
 DR InterPro; IPR000991; GATase_1.
 DR Pfam; PF00117; GATase; 1.
 DR PROSITE; PS00442; GATASE TYPE I; 1.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 201 AA; 22259 MW; D1873B5D78F6BF6B CRC64;

Query Match 14.1%; Score 407; DB 16; Length 201;
 Best Local Similarity 43.0%; Pred. No. 4.8e-20;
 Matches 86; Conservative 38; Mismatches 72; Indels 4; Gaps 4;

QY 28 VVTLLDYGAGNRSVNRNARTLGFIDIKVQKQEDILNAKRLIFPGYGAFAPANDVLRKG 87
 Db 1 MVFIADYGAGNRSVNRNARTLGFIDIKVQKQEDILNAKRLIFPGYGAFAPANDVLRKG 60
 QY 88 LAEALCTYIQNDRPFLGICLGLQLLPESSEBENGPIQGLGLIPGRVGRFESSNGLRVPFH 147

DB 61 FDEAIREHIDKGRSVLIGLGMQLFLSESEMGAYKGLDIVPKVLRFSTSD-KIPQIG 119
QY 148 WHALDIKEGSAILLDDVGNQ-HYFVHSYRANADNKWISSTCSY-GDDPIASIQGNVH 205
DB 120 WNSVDYCKDSVLFNRPVDSQSYFYFVHSYTC-APDEPESVAATFFAGKFKFCSAIEKNGIF 178
QY 206 AVQFHPKSGVGLSLRRF 225
DB 179 AVQFHPKSGVGLSLRRF 198

RESULT 12

Q8FY07 PRELIMINARY; PRT; 261 AA.
AC Q8FY07;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Inidazoleglycerol phosphate synthase, cyclase subunit.
GN HSF OR BR2085.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AE014495; AAN30975.1; -
DR TIGR; BR2085; -
KW Complete proteome.
SQ SEQUENCE 261 AA; 27474 MW; DECA75E3993D1720 CRC64;

Query Match 14.18; Score 405; DB 16; Length 261;
Best Local Similarity 35.18; Pred. No. 9.9e-20;
Matches 110; Conservative 41; Mismatches 104; Indels 58; Gaps 7;
QY 246 LAKRVIACLDVRANDGDLVVTGKDQYDVRERTEENVRNLGKRPVELAGQYILDGDEV 305
DB 3 LKARVIFCLDVK-----DGRVVGKGNF-----VDLIDAGDPVEAARAYDAAGDEL 49
QY 306 FLNITGFRDPLGLPLQLQVLRASENVFPLTVGGGIRFTDANGRYYSLSLEVSEYFR 365
DB 50 FLDTASSD---NRETIFDVVARTAEQCFMPLTVGGGVROVADIR-----KLL 95
QY 366 SGADKVISGDVAVTAIEYKTVGKTSIEQISTVYGQAVVSTDPRVLRKPEDEV 425
DB 96 AGADKVSINTAAVKNP-----FVAEADKFGNQCIIVAID----- 131
QY 426 EFKAIKVSHPGNGEEVAVYQCTVNGREGRPIGAYELAKAVBELGAGEILLNCIDCGQ 485
DB 132 ---AKKVSAGENDR-----WEIFTHGGRQPTGIDAVEFAQKVVDLAGEILLTSMRDGT 184
QY 486 KGKFDILIKLISDAVNPVIASSGAGVADHFSVFNETNASALAAAGIFHRKVEPIKAV 545
DB 185 KAGYDVALTRAVDSVRAPVIASGGVGTLDHLVAGIRGDHATAVLAASIFHFGTYTIGEA 244
QY 546 KEHLKKEGIEVRL 558
DB 245 KRYMAEAGIPMLR 257

RESULT 13

Q8ESS0 PRELIMINARY; PRT; 201 AA.
AC Q8ESS0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Amidotransferase (EC 2.4.2.-).
GN OB0549.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
ridge and its unexpected adaptive capabilities to extreme
environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004594; BAC12505.1; -
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 201 AA; 22321 MW; 36B8F4B049EACAE CRC64;
Query Match 13.8%; Score 396; DB 16; Length 201;
Best Local Similarity 40.2%; Pred. No. 2.7e-19;
Matches 80; Conservative 44; Mismatches 71; Indels 4; Gaps 3;
QY 28 VVTLLDYGNVRSVNRNARTLGFIDKQVQKEDILNARLIPPGVGAAPAMDLVRKG 87
DB 1 MTAIIDYAGNIKSLQFALDKLNKHSIVTTEAAEIKQADSIIILPGVGAFKDAEAIKRLQ 60
QY 88 LAELACTYQNDRPFILGICGLQLFPESSENGPIQGLIGRVRGFRSSNGLRVPFHIG 147
DB 61 LSVIQEEQAKKPLIGICGLQWLFYEQSLDNGWGLGLKGSIKRI--SGEVKVPFHM 118
QY 148 WHALDIKEGSAILLDD-VGNQHYVFSYRANADNKWISSTCSYGDPIASIQGNVH 206
DB 119 WNTLDIQQASPLFDSKLENPVYFVHSYAVSPSENTLLASS-QYGLIPAIQVKGNI 177
QY 207 VQFHPKSGVGLSLRRF 225
DB 178 MQFHPKSGVGLSLRRF 196
RESULT 14
Q939J6 PRELIMINARY; PRT; 201 AA.
AC Q939J6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 22.7 kDa protein.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=81-176;
RX PubMed=11461915;
RA Thibault P., Logan S.M., Kelly J.F., Brisson J.R., Ewing C.P.,
RA Trust T.J., Guerry P.;
RT "Identification of the Carbohydrate Moieties and Glycosylation Motifs
in Campylobacter jejuni Flagellin."
RL J. Biol. Chem. 276:34862-34870(2001).
DR EMBL; AY034084; AAK58487.1; -
DR InterPro; IPR000991; GATase_1.
DR Pfam; PF00117; GATase_1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Hypothetical protein.
SQ SEQUENCE 201 AA; 22732 MW; BC60525EDABC2A3C CRC64;

Query Match 13.7%; Score 394; DB 2; Length 201;
Best Local Similarity 40.5%; Pred. No. 3.7e-19;
Matches 81; Conservative 35; Mismatches 80; Indels 4; Gaps 3;

QY 28 VTLLDYGAGNVRVSRNAIRTLGFDIKDVQKPEDILNAKELIPFGVGAFAPAMDVLRKGL 87
DB 1 MIALIDYKAGNLNSVAKAFKIGAINPIAKNPDKLQKADKLLPGVGSPFKEAMKMLKELG 60

QY 88 LABALCTYI-QNDRPFLGICLGLQLLPESSEENGPIQGLGLIPRVGRPFSSNGLRVPHI 146
DB 61 FIEALKEQVLVQKKPILGICLGLQLPFLERGVEGVCEGLGFIEGVVVKFEDLNLIKPHM 120

QY 147 GWHALDIKEGSAILDVGNQ-HVYFVHSYRANAEDNKEMWISSTCSYGGDDFIASIQGNVH 205
DB 121 GWNELEILKQDPLYQGINNKSDPYFVHSFYVKCKD--EFVSAKAQYGHKEVASLQKDRIP 178

QY 206 AVQFHPKSGVGLSILRRP 225
DB 179 ATQFHPKSNQNLGLKLENF 198

RESULT 15

Q8TS91 PRELIMINARY; PRT; 202 AA.

AC Q8TS91;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Imidazoleglycerol-phosphate synthase, subunit H.
GN HISH OR MA0913.

OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Linton N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR ENBL; AE010756; AAM04346.1; -
DR InterPro; IPR000991; GATase_1.
DR Pfam; PFO0117; GATase; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Complete proteome.
SQ SEQUENCE 202 AA; 21860 MW; 122A2875EBFC371C CRC64;

Query Match 13.7%; Score 393; DB 17; Length 202;
Best Local Similarity 44.3%; Pred. No. 4.4e-19;
Matches 89; Conservative 30; Mismatches 74; Indels 8; Gaps 5;

QY 29 VTLLDYGAGNVRVSRNAIRTLGFDIKDVQKPEDILNAKELIPFGVGAFAPAMDVLRKGL 88
DB 4 IVIDYGLNLSVQKGLHVGANPAISGNPPEILTADGILPVGAFIDAMKCLI--PL 61

QY 89 AEALCTVIQNDRPPLGICLGLQLLPESSEENGPIQGLGLIPRVGRPFSSNGLRVPHIGW 148
DB 62 KGVAIEFAESGKPMGLICLQGVLMSSSEGRLTGGDLIQGRVLRPKSE-LKVPFMGW 120

QY 149 HALDIKEGSAILDVGN-QHVYFVHSYRANAEDNKEMWISSTCSYGGDDFIASI--QKGNVH 205

Db 121 NNIRIKQDHPFLFKGISDGSFVYFVHSY--YVDTTAENTLASCEYGLDFASVSVNSKGNVM 178

QY 206 AVQFHPKSGVGLSILRRFL 226
DB 179 GTQFHPKSGTTGLKILKNFV 199

Search completed: January 20, 2004, 16:27:33
Job time : 43 secs

